STUDIES ON AGGREGATION OF GROUP STRUCTURES AND GROUP ATTRIBUTES THROUGH QUANTIFICATION METHODS

BY

OSAMU KATAI

1979

STUDIES ON AGGREGATION OF GROUP STRUCTURES AND GROUP ATTRIBUTES THROUGH QUANTIFICATION METHODS

A DISSERTATION
PRESENTED TO
KYOTO UNIVERSITY

BY

OSAMU KATAI

FOR THE DEGREE OF
DOCTOR OF ENGINEERING

1979

PREFACE

The present studies have been carried out under the direction of Professor Sousuke Iwai in his laboratory at the Department of Precision Mechanics, Kyoto University. These are collected here as a thesis to be submitted to the Faculty of Engineering, Kyoto University, in fulfillment of requirements for the Degree of Doctor of Engineering.

The studies deal with the efficient ways for aggregation of group structures and group attributes through quantification methods and consist of four parts.

In the first part, the types of data structures prescribing group structures are briefly reviewed; the treatments of the aggregation methods depend on the types of data. The structure of the data prescribing group structures can be broadly categorized into two types: (1) Each member in a group is characterized in relation to a set of attributes. (2) Each member is characterized in relation to other members, i.e., the member-member relationships are characterized.

In the second part, a quantification method for group attributes is developed in relation to discrimination of one subgroup from the others. This quantification method is closely related to Kullback's information theory for statistical decision making and yields an efficient way for aggregation of group attributes.

In the third and the fourth parts, aggregation of group structures which are prescribed by special member-member relationships are considered. The third part deals with the group structures prescribed by Markov transition probabilities amomg members. An extension of Hayashi's quantification method is introduced to quantify each member and its application to aggregation of industrial sectors in inter-industrial relations table is examined.

In the last part, aggregation of sociometric group structures is discussed in relation to the theory of social balance developed by Cartwright and Harary. It is shown that the aggregation is closely related to factor analysis techniques by regarding the adjacency matrices of groups as correlation matrices.

The author would like to express his sincere gratitude to Professor Sousuke Iwai for his supervision and encouragement in the preparation of this thesis.

The aughor is also deeply grateful to Professor Hajime Akashi for his

valuable suggestions, to Professor Sinzō Watanabe and Associate Professor Hidenosuke Nishio of the Faculty of Science, Kyoto University, for their instructive suggestions, to Associate Professor Tomio Kinoshita of the Colledge of Liberal Arts, Kyoto University, for his constructive criticism, and to Professor Jirō Endō of the Colledge of Medical Technology, Kyoto University, and Research Associate Masahiro Fujii of the Faculty of Medicine, Kyoto University, for their valuable comments.

Thanks are also extended to Research Associate Tōru Watanabe, Research Associate Hiromitsu Kumamoto, Research Associate Hiroya Kawai of Osaka University, Lecturer Kazuhide Suhara of Ritsumeikan University, Messrs. Takashi Imanaga, Seizō Takagi and Toshio Tamaru, and all the members of Professor Iwai's laboratory for their helpful suggestions, discussions, and collaborations.

Osamu Katai

Department of Precision Mechanics
Faculty of Engineering
Kyoto University
March, 1979.

CONTENTS

CHAPTER 1 | INTRODUCTION

Aggregations of group structures into subgroup structures provide effective ways for the analyses and the comprehension of group structures, by examining the internal structures of the subgroups and their interrelationships

The premises or the basis of these ways are as follows. Even when the structure of a group has no ordered regularities as a whole, it is sometimes possible that the internal structures of its subgroups have quite ordered one For instance, if a group is composed of two subgroups whose properties are contrary to each other, then its group structure will be regarded as being quite vague or complicated because of its property of being a mixture of thes antipodal properties. However, this difficulty will easily be resolved through the aggregation of the group into the two subgroups.

For example, latent structure analysis is one of the methods for aggregating group structures into subgroups by presuming prerequisite conditions o the internal structures of the subgroups and also on the interrelationships among them (cf. Lazarsfeld[1950 & 1954] and Green[1951]). However, these methods are only applicable to groups with quite restricted structures.

A somewhat more widely available and heuristic way for aggregating group structures, is one using the quantification of each member inside the groups to reflect the internal structures of the groups, and then by aggregating the members with approximately equal quantified values to each other into a subgroup. The quantification means to attach some numerical value to each member. Namely, the method is done through the introduction of topology or proximity into the set of the members in a group, and its applicability to th aggregation problems of group structures depends on the selection of appropri ate quantification methods in accordance with the properties or the types of group structures.

These aggregation methods are important and beneficial for the analyses and treatments of the so-called large scale systems. For instance, in the analysis of economical systems, aggregation problems of industrial sectors ha been studied for the purpose of examining internal structures of economical activities. Their results are useful for predicting future developments, fc making global indices of economical activities, and also for making developmental plannings (cf. Hatanaka[1952] and Ara[1959]). Also, in the field of social psychology, many methods have been devised to analyze the internal

structures of social groups through the aggregation of the group structures into tight subgroups such as cliques. These studies are called sociometric analyses and are particularly beneficial for the comprehension of the personality of individuals inside groups and for improving group behavior (cf. Moreno[1934 & 1960], Ross & Harary [1955], and Tanaka[1975]).

These aggregation problems can be, in the general framework, regarded as the reduction or compression problems of large scale or complex information (data). Data reduction problems are, in some cases, formulated as the reduction problems of the dimensionality of the data. For instance, factor analysis and principal component analysis in data analysis techniques are devised for the data composed of linearly interrelated items. The purpose is to reduce the dimensions of the data, the numbers of the items, and is done by making new items which are linear combinations of the original items (cf. Spearman[1904], Horst[1965], and Anderson[1958]). Also, numerical taxonomy or cluster analysis is one of the reduction methods for the data with nonlinear internal relationships (cf. Sokal & Sneath[1963], Lance & Williams [1967], and Okuno et al.[1971]). Thus, data reduction methods should be, in general, implemented in accordance with the types of the internal structures of the data under investigation.

The structures of the data representing group structures can be categorized broadly into the following two kinds: (1) Each member in a group is characterized in relation to a set of attributes. (2) Each member is characterized in relation to other members, i.e., the member-member relationships are characterized.

This categorization is closely related to the classification of data types introduced by Hayashi (1952) in his theory of data analysis. In the data analysis techniques, each member in a group corresponds to a sample in a population, and each attribute corresponds to an item. His classification is similar to the above categorization except for the supplementation of some additional classifications of items, i.e., he classified the items into two types: those which are concerned with the internal structures of the data and those which prescribe the data in relation to the purposes of the analyses of the data. The items of the latter type are called external criteria (or outside variables). Moreover, he classified these external criteria into those prescribed by numerical values and those prescribed by nominal values, i.e., the items of the first type attach real values to the samples in data and those of the second type give some classifications or aggregations to the set

of the samples in data.

In this framework, he introduced the following four types of data:

I. those prescribed by items in which external criteria are real valued.

II. those prescribed by items in which external criteria are nominal valued.

III. those prescribed by items without external criteria.

IV. those prescribed by internal relationships among samples.

His classification of data types was introduced with the view of developing certain quantification methods, i.e. methods to quantify qualitative data into quantitative data, which will be briefly reviewed in the next chapter. Hence, in the first three types of data, only the qualitative relationship of each sample is prescribed, i.e., whether the sample possesses the attribute or the property represented by the item or not. Also, in the data of the fourth type, what are prescribed as the internal relationships are certain kinds of proximities among the samples (cf. Hayashi[1952] and Saito et al.[1972a]).

Let us consider the aggregation problems of group structures prescribed by the first or the second type data, i.e. prescribed by group attributes and also by external criteria.

In a group structure prescribed by data of the first type, the members inside the group can be allocated in the one-dimensional real axis in terms of the values attached by the external criterion. Hence, its aggregation into subgroups can be easily carried out, without referring to the other items prescribing the internal structure of the group, by the use of the usual methods such as cluster analysis techniques.

In a group structure prescribed by data of the second type, the aggregation of the group is already provided by the external criterion. Hence the aggregation problems of group attributes, i.e. of items, are of primal importance. The aggregation of group attributes means to reduce the number of the attributes. For example, as aforementioned, factor analysis and principal component analysis are the methods to solve this kind of problem for the case of quantitative type of data. However, in this qualitative type of data as mentioned earlier, these methods are, of course, not available. In this case, the aggregation of items means to merge some items into a new item, and the relation of each sample to the new item is modified such that if the sample possesses at least one of the attributes represented by the merged items, then the sample is regarded to possess the attribute represented by the new item. If this is not the case, then the sample is regarded as not possessing the new

attribute. What should be taken into account in this kind of aggregation problem is that the aggregation of items inevitably causes a certain loss of information contained in the original data with respect to the purpose of analysis of the data, i.e. with respect to the aggregation of samples prescribed by the external criterion.

Another kind of representation method for group structures is the method by linear graphs (cf. Harary & Norman[1953] and Harary, Norman, & Cartwright[1965]). This method is frequently and widely used for the group structures prescribed by internal member-member relationships. Namely, the members in a group are represented by vertices and the member-member relationships are represented by lines joining the vertices. Of course, if the relationships do not only prescribe the mere existence or non-existence of member-member relationships but also prescribe their modes, then the groups will be represented as valued graphs. Among the valued graphs, there are two kinds of graphs which are of special interest as the representations of group structures. One consists of those representing Markov chain structures and the other consists of those representing sociometric structures. In the former, the values attached to the lines are in real numbers, not less than 0 and not greater than 1, representing transition probabilities. In the latter, the values are either the sign + or -, which represent the likes-dislikes relationships amomg the members.

For the groups represented by these simple graph structures, the aggregation methods will inevitably be heuristic. As aforementioned, the aggregation method through certain quantification of the members inside groups is one which is widely available for such group structures, by which the members with approximately the same quantified values are aggregated into a subgroup. In the former kind (the groups with Markov chain structures), the ability to specify the apparent subgroups such as ergodic classes can be specified by the above methods is desired (cf. Kemeny & Snell[1963] and Iwahori[1974]). In the latter kind (groups prescribed by sociometric structures), specification by the above methods of the apparent subgroups such as cliques is desired (cf. Festinger[1949], Luce[1950], and Harary[1959]). A clique is a maximal subgroup in which any line joining two members has a positive sign.

As aforementioned, the fourth type in the data classification is related to the prescription of group structures by the internal relationships among members (samples). However, transition probabilities in Markov chain structures have a meaning different to the mere proximities between members (samples) as specified in the data of the fourth type. Furthermore, it is

possible to transform or rewrite the data of the third type into the data prescribing Markov chain structures through identification of items with samples. Therefore, it is to be expected that some extensions of the quantification method for the data of the third type introduced by Hayashi give rise to an appropriate aggregation method for the groups with Markov chain structures.

In the sociometric group structures, what is specified in a member-membe relationship is a likes-dislikes relationship, and can be regarded as a kind of proximity between the members. Hence, it can be expected that the quantification method for the data of the fourth type is available for the aggregation problems of the sociometric group structures. However, some modifica tions should be incorporated so that the apparent subgroups in sociometric structures such as cliques can be detected or specified by the method.

The object of this thesis is to establish some effective aggregation methods for the attributes prescribing group structrues and also for the members in the groups prescribed by Markov chain structures or by sociometric structures by introducing certain quantification methods for the attributes o the members.

In the next chapter, the classification of the data types and the quanti fication methods by Hayashi are briefly reviewed in relation to a more ganera framework, i.e. classification of scales and scaling methods.

In Chapter 3, we develop a quantification and an aggregation method for the attributes of groups whose structures can be described by a special kind of the second type data; we consider the case where the external criteria aggregate the samples into two subgroups, i.e., the sets of the members (samples) are dichotomized. Hayashi's quantification method for the data of the second type is not suited to deal with the aggregation problems of the items, because it is difficult by his method to evaluate the aforementioned information loss due to the aggregation of items. The reason of the introduction of such special restriction into the external criteria is that it mak it possible to develop a quantification method for the items which is closely related to the information loss. The data representing these broadly aggregated group structures, i.e. dichotomized group structures, are inevitably in statistical forms. The relation of each subgroup (specified by the external criteria) to each item will, in general, be represented as the probability th the subgroup has the attribute expressed by the item. In order to develop such a quantification method for the items, certain reasonable prerequisite

conditions on the quantified values are presumed, and it is shown that the quantification method is uniquely determined by the conditions. Then the properties of the quantification method are analyzed in relation to some information theoretic measures and also in relation to the dichotomy given by the external criteria. By the use of the quantification method, an effective aggregation method for the items is developed by examining the relationship between the amount of change of the quantified values due to the aggregation and the above information loss. As an application of the above quantification method, the behavior and the efficiency of the learning model introduced by Luce(1959) are examined.

In Chapter 4, a quantification method which is suited to the aggregation of group structures prescribed by Markov chains is developed. First, the data of the third type are transformed into Markov chain structures by identifying samples (members) with items (group attributes). Some extensions of Hayashi's quantification method for the data of the third type give rise to the above quantification. This quantification method is closely related to the evaluation of dependence between two random variables studied by Gebelein (1941) and Kramer(1960). In order to evaluate the validity of the above quantification method in relation to the aggregation of Markov chain structures, we examine how the qualitative properties of Markov chains such as the classfication of states (ergodic classes, transient classes, cyclic classes etc.) are reflected on the quantified values for the states (members). As a practical application of the quantification method, we consider the aggregation problems of industrial sectors in input-output tables and clarify the structure of the Japanese economical activities.

Chapter 5 deals with the aggregation problems of social groups prescribed by sociometric structures into tight subgroups such as cliques through certain quantification methods. This problem is closely related to the so-called balance theory of social groups introduced by Heider(1946) and Cartwright & Harary(1956). A balanced state of a social group is defined as consisting of one or two cliques. By introducing the notion of sign vectors, it is verified that a modified method of Hayashi's quantification method for the data of the fourth type can specify the one or two cliques in balanced groups. Moreover, it is shown that this method is also available for the derivation of the minimum balancing sets of unbalanced groups by referring to some factor analysis techniques, where a minimum balancing set is the set of interpersonal relations, with the minimum number of elements, whose sign changes yield a balanced state. That is to say, the method gives the most economical way to make un-

balanced groups balanced. In other words, it also gives an aggregation metho for unbalanced social groups into at most two tight subgroups which are nearly cliqual. Another kind of balancing set of practical importance for unbalance social groups is the minimal balancing sets. Some derivation methods for these sets are also considered. In the derivation, some graph theoretical methods are necessarily incorporated to the above method, because of the notio of balance being graph theoretical. Particularly, when the graph structures of social groups are planar, the above graph theoretical methods are quite effective. These notions of balancing sets are closely related to the quantification of the degree of balance of social groups (cf. Abelson & Rosenberg [1958], Harary[1959], and Flament[1963]). In actual social groups, however, their sociometric structures are sometimes inconstant, i.e., the interpersonal relationships inside groups fluctuate between positive and negative. Hence, in this chapter, the degree of balance is considered in a statistical sense, and also some types of social balance are introduced and examined.

CHAPTER 2 | A BRIEF REVIEW OF HAYASHI'S CHARACTERIZATION OF DATA TYPES AND QUANTIFICATION METHODS

2.1 Introduction

In this chapter, Hayashi's classification of data types and his quantification methods are briefly reviewed.

As mentioned in Chapter 1, quantification methods in general can be regarded as certain kinds of scaling methods, i.e., the methods to transform the data prescribed in lower scales to the data in higher scales.

First, we will briefly sketch the classification of scales and their characterization in Section 2.2. Then, in Section 2.3, a brief review of Hayashi's classification of data types and of his quantification methods is given.

2.2 Classification of Scales and Their Characterization

The data analysis techniques such as quantification methods and scaling methods which transform qualitative data into quantitative data are widely used to analyze the phenomena in behavioral, social, natural, and manegement sciences, etc. (cf. Stouffer et al. [1950], Hayashi & Murayama [1969], Takagi [1971 & 1972], Romney, Shepard, & Nerlove [1972], and Nishida & Ara [1976]). The treatment of the data in the techniques varies according to the types of scales used to represent the original data, i.e., the way how the data are measured or obtained. Hence, first of all, we will sketch briefly the classification of the types of scales and their characterization.

2.2.1 Classification of Scales

The most fundamental and famous classification of scales is given by Stevens(1951) as follows(cf. Saito, Ogawa, & Nojima[1972b], and Tanaka[1973]):

(1) Nominal Scale

In the nominal scales, each value represents only the classification of data, i.e., what is shown by nominal scales is the equality and the inequality among data.

(2) Ordinal Scale

In the ordinal scales, the order relationship " > " is defined among the

values, which is a binary relation and is transitive but not symmetric and reflexive, i.e.,

$$\left.\begin{array}{l} \rho > \rho',\ \rho' > \rho'' \rightarrow \rho > \rho'' \\ \rho > \rho' \rightarrow \rho' \not> \rho \\ \rho \not> \rho \end{array}\right\}. \qquad (2.2.$$

(3) Interval Scale

In this kind of scale, not only the order relationship but also the distanc between the values are meaningful. The distance is defined as the differenc between the values.

(4) Logarithmic Interval Scale

In the logarithmic interval scales, the distance between two values is defined as the logarithmic value of the ratio of one value to the other.

(5) Ratio Scale

In each ratio scale, there exists an origin and the distance between two values are defined as the ratio of the distance of one value from the origin to that of the other.

(6) Absolute Scale

An absolute scale is a kind of ratio scale with a prescribed unit for inter vals. Most of the physical measurements are represented by absolute scales.

2.2.2 Characterization of Scales

The six kinds of scales described in 2.2.1 can be characterized in terms of invariant transformations which preserve the information contained in the data. Also, we refer to the statistical quantities which can be defined in each kind of scales (cf. Saito et al.[1972b]).

(1) Nominal Scale

As aforementioned, nominal scales merely represent the classification of values in data, hence the invariant transformations for nominal scales are only one-to-one transformations on $\mathbf{R}^1$. The meaningful statistical quantitie for nominal scales are frequencies, contingency coefficients etc.

(2) Ordinal Scale

The invariant transformations for ordinal scales are monotonic (increasing) transformations on $\mathbb{R}^1$. The statistical quantities which are meaningful in ordinal scales are medians, cumulative frequencies, percentiles, rank correlation coefficient, etc.

(3) Interval Scale

In interval scales, the affine transformations $\rho' = a\rho + b$ (a,b: const., $a > 0$) are invariant transformations. The meaningful statistical quantities are means, variances, covariances, correlation coefficients etc. together with the quantities in (1) and (2).

(4) Logarithmic Interval Scale

The invariant transformations in this case are the power transformations: $\rho' = a\rho^b$ (a, b: consts, $a > 0$, $b > 0$).

(5) Ratio Scale

In ratio scales, invariant transformations are similarity transformations: $\rho' = a\rho$ (a, const., $a > 0$). The meaningful quantities are geometric means, coefficients of variations etc. together with the quantities in (3).

(6) Absolute Scale

The invariant transformation is identity transformation: $\rho' = \rho$.

These six scales constitute a hierarchy as shown in Fig. 2.2.1.

In the figure, absolute, ratio, interval, and logarithmic interval scales are called *metric scales*, and ordinal, nominal scales are called *nonmetric scales*.

2.3 A Brief Review of Hayashi's Classification of Types of Nominal Data and Quantification Methods

In social sciences, it is sometimes necessary to treat nonmetric data, i.e., the data represented by nonmetric scales. On the other hand, the statistical methods to treat nonmetric data, such as nonparametric statistics, are not developed enough to cover such a wide range of phenomena as found in social sciences. This is due to the fact that the information contained in nonmetric scales is much less than that contained in metric scales. One method to avoid such difficulties is to transform nonmetric scales into metric scales. With an insight into the data, we obtain the necessary information or make some hypotheses or models, enabling us to obtain certain essential features of the phenomena represented in the original data. The methods to transform nonmetric scales into metric scales are called *quantification methods* (cf. Saito et al. [1972a]).

Hayashi's quantification methods vary according to the types of the original data. Hence, in Section 2.3.1, his classification of the types of nominal data is briefly sketched, and then, in Section 2.3.2, his methods are briefly reviewed.

2.3.1 An Unified View of Nominal Data

In the previous section, we mentioned the cassification of data in terms of their measuremental scales, i.e., by the measuring methods for data. It should be noted that even when the data are measured by absolute scales such as the measurement of weight or length, each value in the data does not represent the true value but merely represents that the true value is contained in some neighbourhood of the measured value, and hence, the data are, in the strict sense, represented by nominal scales. Also, it should by noted that the statistical techniques developed so far are mainly based on the least mean square methods, i.e., they mainly treat the linear relationships between data. Hence, it is sometimes beneficial to treat the data measured even by absolute scales as being measured by nominal scales, when the relationships between the crude (original) data are of nonlinear types, and then transform the original values into appropriate values such that the relationships are approximately linear (cf. Hayashi & Murayama[1969]). Thus, we could say that the original form of the data is represented by nominal scales. As mentioned previously, the quantification methods are the techniques which transform nominal or ordinal scales into metric scales, i.e., interval, logarithmic interval, ratio, and absolute scales.

An Unified View of Nominal Data

An unified view for the description of the data in nominal scales is given by Hayashi (1952). He called each attribute in the data an *item* and the subclassifications in each item as the *categories* in the item.

Suppose that we have an opinionaire with n items for m respondents. Each respondent offers a response pattern for the categories in the n items. He called each pattern a *sample* and introduced a dummy variable (*response variable*) $\delta_i(j_k)$ to represent the samples (cf. Hayashi[1952], and Saito et al. [1972a]), where

$$\delta_i(j_r) = \begin{cases} 1 & \text{if the } i^{th} \text{ sample responds to the } r^{th} \text{ category in the } j^{th} \text{ item,} \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{for } i = 1, 2, \cdots, m,\ j = 1, 2, \cdots, n,\ r = 1, 2, \cdots, k_j \qquad (2.3.1$$

and m, n, and k_j are the numbers of the samples, the items, and the categories in the j^{th} item, respectively.

The total number of the categories is given by

$$k = \sum_{j=1}^{n} k_j . \tag{2.3.2}$$

The $n \times k$-dimensional matrix

$$D = (\delta_i(j_k)) \tag{2.3.3}$$

is called the *response matrix*(or the *response table*) of the data(cf. Table 2.3.1).

The above mentioned view of data only concerns the way of representation of data. In developing quantification methods, there is another important point of view: for what purpose the analysis of the data should be done. To incorporate the above point of view with the above representation method of data Hayashi[1952] introduced the following method to represent the purposes of the quatification methods or of the analyses of the data. Namely, he considered that the purposes could be represented by a kind of variable called *outside variable*(*external criterion*).

For instance, let us consider the analysis of the result of a election of, say, a governer. The purpose of the analysis is to clarify the effect of the main factors on the result, which will be beneficial for the predictions for the results of other elections of the same kind, i.e., of other governers.

In this case, each sample corresponds to a voter, each item corresponds to one of the main factors, and the outside variable is a dummy variable representing the candidate to whom the voter(sample) casted a vote. For example, if there exist two candidates, say, A and B, then the dummy variable is set as 1 or 0 according to whether the voter cast his vote to A or B. Another way for the analysis of the data will be given by settint the outside cariable as the numbers of votes cast to the candidates. In the former case, the outside variable is represented by nominal scales(nonmetric scales), and in the latter, the outside variable is represented by absolute scales(metric scales).

2.3.2 Classification of the Types of Nominal Data and of the Quantification Methods

In the previous section, we discussed the representation methods of (original) data and those of the purposes of the anayses. The representation methods of data are based on the notions of items and categories. Hence the data represented by the methods are sometimes called *item-category data*. When item-category data are incorporated with the outside variables, then they are repre-

sented by the extended response matrices as shown in Table 2.3.2.

Hayashi(1952) classified the quantification problems according to the typ of the outside variables and those of the response variables as shown in Table 2.3.3, where the cases A, B, and C correspond respectively to the outside vari able being measured by metric scales, by nonmetric scales, and being nonexiste (there being no outside variables). Also, D corresponds to the case where th response variables represent the existence and nonexistence of responses of samples to items(1 corresponds to the existence, and 0 corresponds to the non-existence). E corresponds to the case in which response variables represent the proximities or similarities(measured by some metric scales) among the samples.

Hence, in the quantification method IV, it is assumed that the response matrices are composed of single item, and, the categories and the samples coin cide with each other, i.e., the response matrices are of the form as shown in Table 2.3.4.

In addition to the above, the quantification method II is applicable only to the case where the outside variables are obtained by absolute judgement. When the outside variables are given by comparative judgement, some other scali methods, e.g. Guttman's scalogram analysis, will be used(cf. Guttman[1950]). Also, in the quantification method IV, when a_{ij}'s are measured by nonmetric scales(cf. Abelson & Tukey[1963]), the methods introduced by Shepard(1962 a & b 1966), Kruskal(1964 a & b), Guttman(1968) etc. will be used.

In the sequel, we will briefly sketch the above four quantification metho introduced by Hayashi. (For the application of these methods, refer to Hayash & Murayama[1969], Nishida & Ara[1976], and IBM[1976].) His methods are mainl based on statistical techniques, except for the quantification method IV.

As mentioned previously, quantification methods are the methods to attach a metric value to each category in each item in the response matrices. Let t value attached to the j_r^{th} category in the response matrix(shown by Table 2.3. or 2.3.2) be denoted by $\rho(j_r)$. He considered that the quantified value $\rho(i)$ for the i^{th} sample is given by

$$\rho(i) = \sum_{j=1}^{n} \sum_{r=1}^{k_j} \delta_i(j_r)\, \rho(j_r) \quad \text{for } i = 1, 2, \cdots, n. \tag{2.3.4}$$

The criteria on which the four quantification methods are based are as follows:

(1) the quantification method I.

To maximize the correlation coefficient κ between the outside variables $\{Y_i\}$ ar

the quantified values $\{\rho(i)\}$ for the samples.

(2) the quantification method II.

To maximize the correlation ratio $\eta^2 \triangleq \sigma_b^2 / \sigma^2$, where σ^2 is the variance of $\{\rho(i)\}$ and σ_b^2 is the variance of $\{\rho(i)\}$ between the strata; each of which is given as a set of samples having the same value of the outside variable.

(3) the quantification method III.

In this case, the quantifications for the categories and for the samples are considered separately. Let the quantified value for the i^{th} sample be denoted by $\nu(i)$. Then the criterion is to maximize the correlation coefficient $\kappa_{\rho\nu}$ between the quantified values(weights) $\{\nu(i)\}$ for the samples and quantified values $\{\rho(j_r)\}$ for the categories.

(4) the quantification method IV.

In this case, the data matrices are shown by Table 2.3.4, i.e., the quantification for the samples $\{\rho(i)\}$ is to be considered. The criterion is to maximize

$$Q = - \sum_{i \neq j} a_{ij} (\rho(i) \quad \rho(j))^2 \tag{2.3.5}$$

subject to

$$\sigma^2 = \text{const.} \quad \text{and} \quad \bar{\rho} = 0, \tag{2.3.6}$$

where σ^2 and $\bar{\rho}$ are the variance and the mean of $\{\rho(i)\}$, respectively.

In the above quantification methods, the original data are measured by nominal scales(the treatment for the data measured by ordinal scales is done by disregarding the orders between the categories, i.e., the data are interpreted as nominal data) and the quantified values are given by interval scales.

The quantification methods I, II, and III are related to linear regression analysis(cf. Saito et al.[1972a], Aoyama[1965 a & b], and Williams[1959]), to discriminant analysis and cluster analysis(cf. Saito et al.[1972a], Johnson [1950], Chino[1963], Friedman & Rubin[1967], and Rao[1952]), and to canonical correlation analysis(cf. Saito et al.[1972a], Hotelling[1936], Lancaster[1958], Aoyama[1965 a & b], and Srikantan[1970]) in the statistical multivariate analysis techniques, respectively. The quantification method IV is connected with multidimensional scaling techniques, particularly with Torgerson's scaling(cf. Torgerson[1952], Gower[1966], Takane[1977], and Saito et al.[1973]).

The gists of the above four criteria are: in case (1), the purpose of the quantification is to make the quantified values $\{\rho(i)\}$ for the samples be the best approximation to the outside variables $\{Y_i\}$; in case (2), to make the values $\{\rho(i)\}$ reflect the classification of the samples given by the outside

variable; in case (3), to make the quantified values $\{\rho(j_r)\}$ for the categories and the quantified values $\{\nu(i)\}$ for the samples reflect the internal structure of the data in the sense that the categories(samples) with similar response patterns have approximately the same quantified values, i.e., to make the quantified values reflect the clusters in the categories(samples); in case (4), to make the quantified values $\{\rho(i)\}$ for the samples reflect the clusters in the group of the samples in the sense that the difference between the quantified values $\rho(i)$ and $\rho(j)$ is the best approximation to the dissimilarity between the samples i and j. This is because the criterion (2.3.5) can be rewritten as

$$Q' \triangleq \sum_{i \neq j} \{(a_m - a_{ij}) - (\rho(i) - \rho(j))\}^2 \to \min, \tag{2.3.5'}$$

where a_m is the maximum value of $\{a_{ij}\}$, i.e., $a_m = \max_{i,j} a_{ij}$. In the above, the quantity $a_m - a_{ij} (> 0)$ can be interpreted as dissimilarity or contra-proximity between the i^{th} sample and j^{th} sample.

In the terminology of scaling methods, the above four types of quantification methods are sometimes called optimal scaling, canonical scaling, or appropriate scoring(cf. Nishisato[1975]).

The solution for each case is as follows(for details, refer to Saito et al [1972a] and Yasuda[1975]):

(1) the quantification method I.

In case (1), the correlation coefficient is given as

$$\kappa = \frac{\frac{1}{m}\sum_{i=1}^{m}(Y_i - \bar{Y})(\rho(i) - \bar{\rho})}{\sigma_Y \cdot \sigma}, \tag{2.3.7}$$

$$\sigma_Y^2 \triangleq \frac{1}{m}\sum_{i=1}^{m}(Y_i - \bar{Y})^2, \tag{2.3.8}$$

$$\sigma^2 \triangleq \frac{1}{m}\sum_{i=1}^{m}(\rho(i) - \bar{\rho})^2, \tag{2.3.9}$$

$$\bar{Y} \triangleq \frac{1}{m}\sum_{i=1}^{m} Y_i, \tag{2.3.10}$$

$$\bar{\rho} \triangleq \frac{1}{m}\sum_{i=1}^{m} \rho(i). \tag{2.3.11}$$

By using the similar methods in linear regression analysis, the criterion yield the next linear equation

$$F \rho \quad {}^t D \mathbf{Y}, \tag{2.3.12}$$

where

$$\mathfrak{p} \triangleq {}^{t}(\rho(1_1), \rho(1_2), \cdots, \rho(1_{k_1}), \cdots, \rho(j_1), \cdots, \rho(j_r), \cdots, \rho(j_{k_j}), \cdots, \rho(n_1), \cdots, \rho(n_{k_n})), \tag{2.3.13}$$

$$\mathbf{Y} \triangleq {}^{t}(Y_1, Y_2, \cdots, Y_m), \tag{2.3.14}$$

$$F \triangleq {}^{t}D\,D = (f(j_r, u_v)), \tag{2.3.15}$$

$$f(j_r, u_v) \triangleq \sum_{i=1}^{m} \delta_i(j_r)\,\delta_i(u_v), \text{ for } r = 1, 2, \cdots, k_j,\ j = 1, 2, \cdots, n,$$
$$v = 1, 2, \cdots, k_u,\ u = 1, 2, \cdots, n, \tag{2.3.16}$$

and D is the response matrix of the data. The solution vector $\mathfrak{p}$ is given by solving the above normal equation (2.3.12).

(2) the quantification method II.

In this case, the outside variable gives the stratification or the classification of the samples. Let the number of the strata be w, the number of the samples in the s^{th} strata be m_s, and i(s) denote the i^{th} sample in the s^{th} stratum. Then the correlation ratio η^2 is given as

$$\eta^2 = \frac{\sigma_b^2}{\sigma^2}, \tag{2.3.17}$$

where

$$\sigma_b^2 \triangleq \sum_{s=1}^{w} \frac{m_s}{m} (\bar{\rho}_s - \bar{\rho})^2, \tag{2.3.18}$$

$$\bar{\rho}_s \triangleq \frac{1}{m_s} \sum_{i(s)=1}^{m_s} \rho(i(s)). \tag{2.3.19}$$

The criterion yields the next equation.

$$H\,\mathfrak{p} = \eta^2\, F'\,\mathfrak{p}, \tag{2.3.20}$$

where

$$F' \triangleq F \quad (r(j_r, u_v)), \tag{2.3.21}$$

$$r(j_r, u_v) \triangleq \frac{1}{m}\, m_{j_r} \cdot m_{u_v}, \tag{2.3.22}$$

$$m_{j_r} \triangleq \sum_{i=1}^{m} \delta_i(j_r), \tag{2.3.23}$$

$$H \overset{\Delta}{=} (g(j_r, u_v) \quad r(j_r, u_v)), \tag{2.3.24}$$

$$g(j_r, u_v) \overset{\Delta}{=} \sum_{s=1}^{w} \frac{g^s(j_r)\, g^s(u_v)}{m_s}, \tag{2.3.25}$$

$$g^s(j_r) \overset{\Delta}{=} \sum_{i(s)=1}^{m_s} \delta_{i(s)}(j_r). \tag{2.3.26}$$

When the matrix F' is regular, then the solution vector ρ is given by the following eigenvalue-eigenvector equation.

$$F'^{-1} H \rho = \eta^2 \rho. \tag{2.3.27}$$

(3) the quantification method Ⅲ.

In this case, the correlation coefficient $\kappa_{\rho\nu}$ between the quantified values $\{\nu(i)\}$ for the samples and those for the categories, i.e. $\{\rho(j_r)\}$, is given by

$$\kappa_{\rho\nu} = \frac{\sigma_{\rho\nu}}{\sigma_\nu \cdot \sigma_\rho}, \tag{2.3.28}$$

$$\sigma_\rho^2 \overset{\Delta}{=} \frac{1}{q} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{k_j} (\rho(j_r) \quad \tilde{\rho})^2 \cdot \delta_i(j_r), \tag{2.3.29}$$

$$\tilde{\rho} \overset{\Delta}{=} \frac{1}{q} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{k_j} \rho(j_r) \cdot \delta_i(j_r), \tag{2.3.30}$$

$$q \overset{\Delta}{=} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{k_j} \delta_i(j_r), \tag{2.3.31}$$

$$\sigma_\nu^2 \overset{\Delta}{=} \frac{1}{q} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{k_j} (\nu(i) \quad \tilde{\nu})^2 \cdot \delta_i(j_r), \tag{2.3.32}$$

$$\tilde{\nu} \overset{\Delta}{=} \frac{1}{q} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{k_j} \nu(i) \cdot \delta_i(j_r), \tag{2.3.33}$$

$$\sigma_{\rho\nu} \overset{\Delta}{=} \frac{1}{q} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{k_j} (\nu(i) \quad \tilde{\nu})(\rho(j_r) \quad \tilde{\rho}) \cdot \delta_i(j_r). \tag{2.3.34}$$

Without loss of generality, we can assume that

$$\tilde{\rho} = \tilde{\nu} = 0. \tag{2.3.35}$$

Then, the criterion is reduced to

$$\begin{cases} \sum_{j_r} \rho(j_r) \dfrac{\delta_i(j_r)}{\sum_{u_v} \delta_i(u_v)} & \lambda\, \nu(i) \quad \text{for } i = 1, 2, \cdots, m, \\ \sum_{i} \nu(i) \dfrac{\delta_i(j_r)}{\sum_{t} \delta_t(j_r)} & = \mu\, \rho(j_r) \quad \text{for } r = 1, 2, \cdots, k_j,\ j = 1, 2, \cdots, n. \end{cases} \tag{2.3.36}$$

The above equations are also reduced to the following eigenvalue-eigenvector equations:

$$\begin{cases} \sum_{u} \nu(u) \left(\sum_{j_r} \dfrac{\delta_i(j_r)\, \delta_u(j_r)}{\sum_{p_q} \delta_i(p_q) \cdot \sum_{t} \delta_t(j_r)} \right) & \kappa_{\rho\nu}^{\,2}\, \nu(i) \\ & \quad \text{for } i = 1, 2, \cdots, m, \\ \sum_{v_w} \rho(v_w) \left(\sum_{i} \dfrac{\delta_i(j_r)\, \delta_i(v_w)}{\sum_{u} \delta_u(j_r) \cdot \sum_{c_d} \delta_i(c_d)} \right) & = \kappa_{\rho\nu}^{\,2}\, \rho(j_r) \\ & \quad \text{for } r = 1, 2, \cdots, k_j,\ j = 1, 2, \cdots, n. \end{cases} \tag{2.3.37}$$

The optimum quantified values $\{\rho(j_r)\}$ and $\{\nu(i)\}$ are given by solving the above equations.

(4) the quantification method IV.

The criterion Q defined by (2.3.5) can be rewritten as follows:

$$Q = \sum_{i} \sum_{j} b_{ij}\, \rho_i\, \rho_j, \tag{2.3.38}$$

where

$$\begin{aligned} b_{ij} &= a_{ij} + a_{ji} \quad \text{for } i \neq j, \\ b_{ii} &= \sum_{\substack{j=1 \\ j \neq i}}^{m} (a_{ij} + a_{ji}) \quad \text{for } i = 1, 2, \cdots, m. \end{aligned} \tag{2.3.39}$$

Namely, the criterion Q is given as a quadratic form for the symmetric matrix $B = (b_{ij})$. As it is wellknown in the theory of linear algebra, the optimum quantification vector $\boldsymbol{\rho}$ is given as the eivenvector of B associated with the maximum eigenvalue, i.e.,

$$B\boldsymbol{\rho} = \lambda^* \boldsymbol{\rho}, \tag{2.3.40}$$

where

$$\rho \triangleq {}^t(\rho(1),\ \rho(2),\ \cdots,\ \rho(m)). \tag{2.3.41}$$

and λ^* is the maximum eigenvalue of the symmetric matrix B.

In response matrices, it sometimes occurs that some samples have the same response pattern and some categories in an item have the same response value fo each sample. For instance, in the response matrix shown by Table 2.3.5, the second and the third samples and also the fifth and the seventh samles have the same response patterns, respectively. Also, categories 2_2 and 2_4 in the second item share the same response values for all of the samples, i.e., $\delta_i(2_2) = \delta_i(2_4)$ for i 1, 2, $\cdots$, 7. In such cases, we can aggregate the classifications of the samples and of the categories into a smaller number of classes as shown in Table 2.3.6(cf. Saito et al.[1972a]). In the table, the second sampl corresponds to the second and the third samples in the original matrix shown by Table 2.3.5. Also, the fourth sample corresponds to the fifth and the seventh samples, and the category 2_2 corresponds to the categories 2_2 and 2_4 in the original matrix. Thus, in the new response matrix, each sample and each category represent the types of response patterns. Hence they are called a *type o sample*(or a *sample type*) and a *type of category*(or a *category type*), respectivel

In general, the new response matrix is represented by the table shown in Table 2.3.7, where $\varepsilon_i(j_r)$ is the dummy variable representing the response value of the i^{th} sample type to the r^{th} category type in the j^{th} item i.e.,

$$\varepsilon_i(j_r) = \begin{cases} 1 & \text{if the } i^{th} \text{ sample type responds to the } r^{th} \text{ category type in the } j^{th} \text{ item,} \\ 0 & \text{otherwise,} \end{cases} \tag{2.3.42}$$

and where f_i is the number of the samples in the i^{th} sample type, and g_{j_r} is that of the categories in the r^{th} category type in the j^{th} item for i = 1, 2, $\cdots$, n, r 1, 2, $\cdots$, k_j, and j - 1, 2, $\cdots$, n.

The above four quantification methods can also be carried out in terms of the above new response matrix(cf. Saito et al.[1972a]). For instance, the quantification method III based on the new matrix can be stated by the following equations (2.3.36') and (2.3.37') which correspond to (2.3.36) and (2.3.37), respectively.

$$\begin{cases} \sum\limits_{j_r} \rho(j_r) \dfrac{g_{j_r}\, \varepsilon_i(j_r)}{\sum\limits_{u_v} g_{u_v}\, \varepsilon_i(u_v)} = \lambda\, \nu(i) \\ \qquad\qquad \text{for } i = 1, 2, \cdots, m, \end{cases}$$

$$\left\{ \sum_i \nu(i) \frac{f_i \varepsilon_i(j_r)}{\sum_t f_t \varepsilon_t(j_r)} = \mu \rho(j_r) \right.$$
$$\text{for } r = 1, 2, \cdots, k_j,\ j = 1, 2, \cdots, n. \quad (2.3.36')$$

$$\begin{cases} \sum_u \nu(u) \left(\sum_{j_r} \frac{g_{j_r} \varepsilon_i(j_r)}{\sum_{p_q} g_{p_q} \varepsilon_i(p_q)} \cdot \frac{f_u \varepsilon_u(j_r)}{\sum_t f_t \varepsilon_t(j_r)} \right) = \kappa_{\rho\nu}^2 \nu(i) \\ \qquad \text{for } i = 1, 2, \cdots, m, \\ \sum_{v_w} \rho(v_w) \left(\sum_i \frac{f_i \varepsilon_i(j_r)}{\sum_u f_u \varepsilon_u(j_r)} \cdot \frac{g_{v_w} \varepsilon_i(v_w)}{\sum_{c_d} g_{c_d} \varepsilon_i(c_d)} \right) = \kappa_{\rho\nu}^2 \rho(j_r) \\ \qquad \text{for } r = 1, 2, \cdots, k_j, \\ \qquad\qquad j = 1, 2, \cdots, n. \end{cases} \quad (2.3.37')$$

In the above, $\nu(i)$ and $\rho(j_r)$ are the quantified values for the i^{th} sample type and r^{th} category type in the i^{th} item, respectively. The above modification of the response matrices does not change the quantified values: if the i^{th} sample in the original response matrix is contained in the i'^{th} sample type in the new response matrix, then we have $\nu(i) = \nu(i')$, and this also holds for the categories.

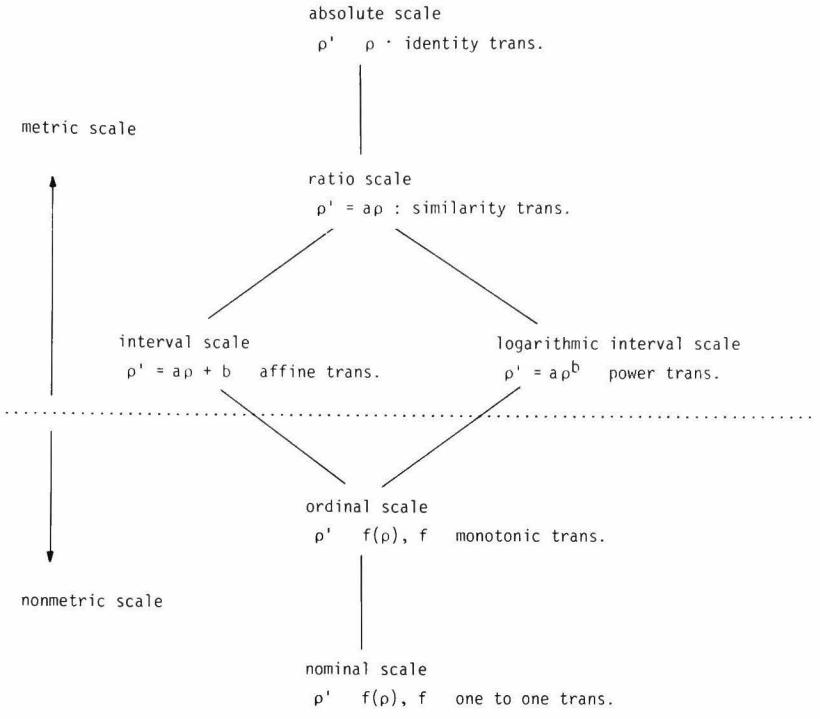


Fig. 2.2.1. Hierarchy among six kinds of scales.

item / category / sample	1		....		....	j					....	n			
	l_1	l_2	....	l_{k_1}	....	j_1	....	j_r	....	j_{k_j}	....	n_1	n_2	....	n_{k_n}
1	1	0	....	1		$\delta_1(j_1)$	....	$\delta_1(j_r)$	....	$\delta_1(j_{k_j})$		0	1	....	1
2	1	1	....	1		$\delta_2(j_1)$	....	$\delta_2(j_r)$	....	$\delta_2(j_{k_j})$		1	1	....	0
													.		
													.		
												.	.		
i	0	0	....	0		$\delta_i(j_1)$	....	$\delta_i(j_r)$	....	$\delta_i(j_{k_j})$		0	1	....	0
.	.							.					.		
	.												.		
	.	.				.							.		
m	1	1	....	1		$\delta_m(j_1)$	....	$\delta_m(j_r)$	....	$\delta_m(j_{k_j})$		1	0	....	1

Table 2.3.1. An example of response matrices composed of m samples and n items.

outside variable	item / category / sample	1	.....	j	.....	n
		$l_1 \cdots\cdots k_1$	.....	$j_1 \cdots\cdots j_r \cdots\cdots j_{k_j}$	.....	$n_1 \cdots\cdots n_{k_n}$
Y_1	1	$\delta_1(l_1)\cdots \delta_1(k_1)$		$\delta_1(j_1)\cdots \delta_1(j_r)\cdots \delta_1(j_{k_j})$		$\delta_1(n_1)\cdots \delta_1(n_{k_n})$
Y_2	2	$\delta_2(l_1)\cdots \delta_2(k_1)$		$\delta_2(j_1)\cdots \delta_2(j_r)\cdots \delta_2(j_{k_j})$		$\delta_2(n_1)\cdots \delta_2(n_{k_n})$
.		.				
	.				.	
.			.			
Y_i	i	$\delta_i(l_1)\cdots \delta_i(k_1)$	.	$\delta_i(j_1)\cdots \delta_i(j_r)\cdots \delta_i(j_{k_j})$		$\delta_i(n_1)\cdots \delta_i(n_{k_n})$
.		.			.	
			.			
						.
Y_m	m	$\delta_m(l_1)\cdots \delta_m(k_1)$		$\delta_m(j_1)\cdots \delta_m(j_r)\cdots \delta_m(j_{k_j})$	.	$\delta_m(n_1)\cdots \delta_m(n_{k_n})$

Table 2.3.2. The extended response matrix, where Y_i denotes the value of external criterion for i^{th} sample.

classification of the quantification methods	type of outside variables	type of response variables
quantification method I	A	D
quantification method II	B	D
quantification method III	C	D
quantification method IV	C	E

Table 2.3.3. Classification of Hayashi's quantification methods.

sample (type)	1	2	.....	j	.....	n
1	a_{11}	a_{12}	.....	a_{1j}	.....	a_{1n}
2	a_{21}	a_{22}	.....	a_{2j}	.....	a_{2n}
⋮	⋮	⋮		⋮		⋮
i	a_{i1}	a_{i2}	.....	a_{ij}	.....	a_{in}
⋮	⋮	⋮		⋮		⋮
n	a_{n1}	a_{n2}	.....	a_{nj}	.....	a_{nn}

Table 2.3.4. The type of response matrices for the data analyzed by the quantification method IV, where a_{ij} represents the proximity of the i^{th} sample to the j^{th} sample.

item / category / sample	1			2			
	1_1	1_2	1_3	2_1	2_2	2_3	2_4
1	1	0	0	0	1	0	1
2	1	1	0	1	1	1	1
3	1	1	0	1	1	1	1
4	0	1	0	1	1	1	1
5	0	0	1	0	0	1	0
6	1	1	1	0	1	0	1
7	0	0	1	0	0	1	0

Table 2.3.5. An example of response matrices which can be aggregated into those with smaller numbers of samples and categories.

item / category type / sample type	1			2		
	1_1	1_2	1_3	2_1	2_2	2_3
1	1	0	0	0	1	0
2	1	1	0	1	1	1
3	0	1	0	1	1	1
4	0	0	1	0	0	1
5	1	1	1	0	1	0

Table. 2.3.6. The aggregated response matrix.

item / category type / sample type	1			···	j					···	n			number of sample
	1_1	···	1_{k_1}	···	j_1	···	j_r	···	j_{k_j}	···	n_1	···	n_{k_n}	
1	$\varepsilon_1(1_1)$	···	$\varepsilon_1(1_{k_1})$		$\varepsilon_1(j_1)$	···	$\varepsilon_1(j_r)$	···	$\varepsilon_1(j_{k_j})$	·	$\varepsilon_1(n_1)$	···	$\varepsilon_1(n_{k_n})$	f_1
2	$\varepsilon_2(1_1)$	···	$\varepsilon_2(1_{k_1})$		$\varepsilon_2(j_1)$	···	$\varepsilon_2(j_r)$	···	$\varepsilon_2(j_{k_j})$		$\varepsilon_2(n_1)$	···	$\varepsilon_2(n_{k_n})$	f_2
·	·				·					·				·
·					·				·	·	·		·	·
							·			·	·		·	·
i	$\varepsilon_i(1_1)$		$\varepsilon_i(1_{k_1})$	·	$\varepsilon_i(j_1)$		$\varepsilon_i(j_r)$		$\varepsilon_i(j_{k_j})$		$\varepsilon_i(n_1)$		$\varepsilon_i(n_{k_n})$	f_i
	·				·		·			·			·	·
	·				·						·		·	·
·	·		·		·					·	·		·	·
·				·			·		·		·		·	·
m	$\varepsilon_m(1_1)$		$\varepsilon_m(1_{k_1})$	·	$\varepsilon_m(j_1)$		$\varepsilon_m(j_r)$		$\varepsilon_m(j_{k_j})$		$\varepsilon_m(n_1)$		$\varepsilon_m(n_{k_n})$	f_m
number of categories	g_{1_1}	····	$g_{1_{k_1}}$	···	g_{j_1}	···	g_{j_r}	···	$g_{j_{k_j}}$	···	g_{n_1}	···	$g_{n_{k_n}}$	

Table 2.3.7. Response matrix consisted of sample types and category types.

CHAPTER 3 | INTRODUCTION OF A QUANTIFICATION METHOD FOR ITEMS AND CATEGORIES IN GROUP ATTRIBUTES AND COSIDERATIONS ON ITEM-CATEGORY AGGREGATION PROBLEMS

3.1 Introduction

As mentioned in Chapter 1, in a group structure prescribed by data of the second type, the aggregation of members(samples) in the group is already given by the external citerion. Hence the aggregation of items and categories in group attributes is of primal importance.

Also, in this kind of group structure, quantification of categories and items should be made so that the quantified values reflect the aggregation of members given by the external criteria, because the purpose of the analyses of these group structures, which are expressed by the external criteria, are to specify the membership of each member(sample) by the use of group attributes. For instance, in the quantification method by Hayashi for the data of the second type, which was briefly reviewed in Section 2.3.2, the correlation ratio is used to evaluate the effectiveness of the quantified values in specifying the memberships.

As mentioned in Chapter 1, the aggregation of group attributes inevitably causes certain loss of information contained in the attributes for specifying the memberships. Hence the aggregation should be done so that the information loss is made as little as possible.

In this chapter, we deal with the aggregation problems of attributes of groups whose structures are described by a special kind of second type data. Namely, the members(samples) of the groups are dichotomized by the external criteria. The reason for the introduction of such a special restriction into the external criteria is that it makes it possible to develop a quantification method for items and categories which is closely related to the information loss. The aforementioned Hayashi's quantification method is available for the aggregation of group attributes in general cases, without imposing such restriction on the external criteria. However, the evaluation expressed by the correlation ratio is hard to deal with in relation to the above information loss Namely, it is difficult to evaluate the decrease of the correlation ratio due to the aggregation of items and categories.

For these broadly aggregated group structures, i.e. dichotomized group structures, the corresponding response matrices are inevitably in statistical

forms. Hence this quantification problem of items and categories is closely related to that of items and categories in contingency tables.

For the quantification problems of group attributes, in genetal, the quantification methods based on the information theory are not sufficiently developed so far. However, these methods are expected to be suited for dealing with these kinds of group structures, because the response matrices corresponding to these group structures are in statistical forms, and also it is necessary, as aforementioned, to evaluate the above information loss due to the aggregation of items and categories.

In Section 3.2, we set up certain reasonable criteria for the quantification ρ for each category in each item and show that ρ is uniquely determined by the above criteria. According to the quantification ρ, we introduce a quantification for each item in such response matrices, which represents the effectiveness of the item for discrimination between the two subgroups prescribed by the external criteria. In Section 3.3, the properties of the quantification for items, which we call mean information intensity, are analyzed with reference to some information theoretic measures, and their relation to the discrimination rates of likelihood ratio tests is examined. In Section 3.4, we consider an efficient way of item-category aggregation in response matrices based on the quantification introduced in Section 3.2. Furthermore, we apply the above method to the quantization problems for continuous measurements. In Section 3.5, we apply the above quantification method to the analyses of the learning behavior of Luce's beta model, by introducing a conditional probability learning model whose behavior is eqivalent to the Luce's model.

3.2 Introduction of a Quantification Method for Items and Categories and of a Measure of Discriminability between Subgroups

In this section, we consider the quantification of items and categories for response matrices in which the number of sample types are two and the response variables represent the response probabilities of sample types to category types.

Let R and $\overline{R}$ be two sample types with *a priori* probabilities P(R) and $P(\overline{R})$. Also, let $X_1, X_2, \cdots$, and X_n be the items, and $P(X_i/R)$ and $P(X_i/\overline{R})$ be the conditional probabilities of the item X_i under R and $\overline{R}$, respectively. Then the response matrices are shown by Table 3.2.1, where $p_j^i = P(X_i = i_j/R)$ and $q_j^i = P(X_i = i_j/\overline{R})$.

Under the above probability structures, we consider the quantification of

category types 1_1, 1_2, $\cdots$, and n_{k_n} from the following point of view.

In the quantification method in Section 2.3.2, each of the category types is evaluated by referrring to all the probabilities in the table. However, the table says nothing about the stochastic dependence between items. So, we assume that items X_1, X_2, $\cdots$, and X_n are stochastically independent. Thus, each category type i_j must be quantified by referring to p^i_j and q^i_j, i.e.,

$$\rho(i_j) = F^i_j(p^i_j ; q^i_j). \tag{3.2.1}$$

Assuming the symmetry in the ordering of the items and also that of the category types, $\rho(i_j)$ must be described as follows:

$$\rho(i_j) = F(p^i_j ; q^i_j) \quad \text{for } j = 1, 2, \cdots, k_i, \; i = 1, 2, \cdots, n. \tag{3.2.2}$$

Also, assuming the symmetry in the sample types R and $\bar{R}$, we have

$$F(p^i_j ; q^i_j) = F(q^i_j ; p^i_j). \tag{3.2.3}$$

We suppose that function F has the following form:

$$F(x ; y) = G(x) \quad G(y). \tag{3.2.4}$$

From the requirement that the quantified value ρ should reflect the distinction between R and $\bar{R}$, as mentioned in Section 3.1, we impose the next condition on ρ.

$$S_R(X_i) \triangleq E_{X_i}[\rho(X_i)\cdot d(X_i)] \geq 0 \quad \text{for an arbitrary probability structure of } X_i, \tag{3.2.5}$$

where $d(X_i)$ is a dummy variable representing the distinction between the sample types as follows:

$$d(X_i) \quad \begin{cases} 1 & \text{if } X_i \in R, \\ -1 & \text{if } X_i \in \bar{R}. \end{cases} \tag{3.2.6}$$

From (3.2.4), (3.2.5), and (3.2.6), we obtain

$$P(R) \sum_{j=1}^{k_i} p^i_j \{ G(p^i_j) \quad G(q^i_j)\} + P(\bar{R}) \sum_{j=1}^{k_i} q^i_j \{ G(q^i_j) \quad G(p^i_j)\} \geq 0$$

for arbitrary probabilities $p^i_1, p^i_2, \cdots, p^i_{k_i}, q^i_1, q^i_2, \cdots, q^i_{k_i}$, $P(R)$, and $P(\bar{R})$. (3.2.7)

The above condition is equivalent to

$$\sum_{j=1}^{k_i} p_j^i \{G(p_j^i) - G(q_j^i)\} \geq 0, \text{ and } \sum_{j=1}^{k_i} q_j^i \{G(q_j^i) - G(p_j^i)\} \geq 0,$$

for arbitrary probabilities $p_1^i, p_2^i, \cdots, p_{k_i}^i, q_1^i, q_2^i, \cdots, q_{k_i}^i$.

(3.2.8)

Hence, in general, function G must satisfy the following(cf. Hendrickson & Buehler[1971]):

$$\sum_j a_j G(a_j) \geq \sum_j a_j G(b_j) \quad \text{for arbitrary } a_1, a_2, \cdots, a_r, b_1, b_2, \cdots, b_r$$

such that $a_j \geq 0$, $b_j \geq 0$ for $j = 1, 2, \cdots, r$

and $\sum_j a_j = \sum_j b_j = 1$. (3.2.9)

THEOREM 3.2.1 : If function G is C^1-class and satisfies (3.2.9) and is not constant, then G is represented by the next formula:

$$G(p) = c \cdot \ln p + d, \tag{3.2.10}$$

where c and d are arbitrary constants such that

$$c > 0. \tag{3.2.11}$$

PROOF : When a_j's are regarded as constants and b_j's are regarded as variables, the function $\sum_j a_j G(b_j)$ has its maximum value at $b_j = a_j$ for $j = 1, 2, \cdots, r$. Using Lagrange multiplier, we have

$$a_j \cdot G'(a_j) = c (> 0), \tag{3.2.12}$$

it follows that

$$G'(a_j) = \frac{c}{a_j}. \tag{3.2.13}$$

This leads to (3.2.10). □

From the standpoint of quantification methods, the constants c and d have no essential meaning and can be set as c = 1 and d = 0, i.e.,

$$G(p) = \ln p. \tag{3.2.14}$$

Hence we obtain

$$\rho(i_j) = \ln \frac{p_j^i}{q_j^i} = \ln \frac{P(X_i = i_j / R)}{P(X_i = i_j / \bar{R})}, \tag{3.2.15}$$

, i.e., $\rho(i_j)$ is exactly the same as the logarithmic likelihood ratio of the event $X_i = i_j$.

Kullback(1959) considered the evaluation of information content contained in a sample value of a random variable for discriminating one hypothsis from the other. In his terminology, the above quantity $\rho(i_j)$ is called the information in the event $X_i = i_j$ for discrimination in favor of hypotheses R against $\bar{R}$, by regarding the sample types R and $\bar{R}$ as hypotheses.

The measure defined by (3.2.5) is given as follows:

$$\begin{aligned} S_R(X_i) &= \sum_{j=1}^{k_i} \{P(R)\cdot P(X_i = i_j / R)\cdot\rho(i_j) \quad P(\bar{R})\cdot P(X_i = i_j / \bar{R})\cdot\rho(i_j)\} \\ &= \sum_{j=1}^{k_i} \{P(R)\ p_j^i \ln\frac{p_j^i}{q_j^i} \quad P(\bar{R})\ q_j^i \ln\frac{p_j^i}{q_j^i}\} \\ &= \sum_{j=1}^{k_i} \{P(R)\ p_j^i \ln\frac{p_j^i}{q_j^i} + P(\bar{R})\ q_j^i \ln\frac{q_j^i}{p_j^i}\}. \end{aligned} \tag{3.2.16}$$

The above measure represents the degree of separation between the sample types R and $\bar{R}$ with respect to item X_i.

In the sequel, we consider the relation of the above quantification method to the statistical decision methods. Let $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be a sample value of $\mathbf{X}$. Also, let us assume the following additivity for the quantification of x_i's. Namely, we assume that the quantification of $\mathbf{x}$ is given as

$$\rho(\mathbf{x}) = \sum_{i=1}^{n} \rho(x_i) \qquad \sum_{i=1}^{n} \ln\frac{P(X_i = x_i / R)}{P(X_i = x_i / \bar{R})}. \tag{3.2.17}$$

If the random variables $X_1, X_2, \cdots$, and X_n are, as aforementioned, stochastically independent of each other within each sample type (R or $\bar{R}$), then $\rho(\mathbf{x})$ is

$$\rho(\mathbf{x}) = \ln\frac{P(\mathbf{X} = \mathbf{x} / R)}{P(\mathbf{X} = \mathbf{x} / \bar{R})}. \tag{3.2.18}$$

Namely, ρ is the logarithmic likelihood ratio of the sample value $\mathbf{x}$ of $\mathbf{x}$. From this, the quantification ρ can also be used in a statistical decision making such as the Bayesian or Neyman-Pearson tests of hypotheses, for both tests can be reduced to the so-called likelihood ratio tests, i.e., the belongingness of a sample value $\mathbf{x}$ to sample type R or $\bar{R}$ is decided according to whether $\rho(\mathbf{x})$ is larger than a certain critical value or not. Hence, $S_R(\mathbf{X}) \triangleq E[\rho(\mathbf{X})\cdot d(\mathbf{X})]$ has certain relationships with discrimination rates in a Bayesian or Neyman-

Pearson's sense(or, in general, in likelihood ratio tests).

Let k be the critical value of a likelihood ratio test and s(t;k) be as follows:

$$s(t;k) \quad \begin{cases} 1 & \text{if } t > \ln k, \\ -1 & \text{if } t < \ln k. \end{cases} \tag{3.2.19}$$

Then the discrimination rate P_c of the test is described by

$$P_c = \frac{1}{2} + \frac{1}{2} E_{\mathbf{X}}[s(\rho(\mathbf{X});k)\cdot d(\mathbf{X})]. \tag{3.2.20}$$

The function s being a nonlinear threshold function, the evaluation of P_c is, in general, difficult. The measure $S_R(\mathbf{X}) \quad E_{\mathbf{X}}[\rho(\mathbf{X})\cdot d(\mathbf{X})]$ has the following useful property and can be regarded as a practical approximation of P_c. We call the measure *mean information intensity of* $\mathbf{X}$ *with respect to the sample types* $R = \{R, \bar{R}\}$.

THEOREM 3.2.2 : Let $(\mathbf{X}, \mathbf{Y})$ denote the joint vector of $\mathbf{X}$ and $\mathbf{Y}$. Then we have

$$S_R((\mathbf{X}, \mathbf{Y})) = S_R(\mathbf{X}) + S_R(\mathbf{Y} \mid \mathbf{X}), \tag{3.2.21}$$

where $S_R(\mathbf{Y} \mid \mathbf{X})$ is defined as follows:

$$S_R(\mathbf{Y} \mid \mathbf{X}) \triangleq E_{\mathbf{X}}[\, E_{\mathbf{Y}}[\, \rho(\mathbf{Y} \mid \mathbf{X})\cdot d((\mathbf{X}, \mathbf{Y})) \mid \mathbf{X}]], \tag{3.2.22}$$

$$\rho(\mathbf{Y} \mid \mathbf{X}) \triangleq \ln \frac{P(\mathbf{Y} / \mathbf{X}, R)}{P(\mathbf{Y} / \mathbf{X}, \bar{R})}. \tag{3.2.23}$$

Particularly, if $\mathbf{X}$ and $\mathbf{Y}$ are stochastically independent in each sample type, then

$$S_R((\mathbf{X}, \mathbf{Y})) = S_R(\mathbf{X}) + S_R(\mathbf{Y}). \tag{3.2.24}$$

PROOF : From the definition of S_R, we have

$$S_R((\mathbf{X}, \mathbf{Y})) = E_{(\mathbf{X}, \mathbf{Y})}\left[\, \ln \frac{P(\mathbf{X}, \mathbf{Y} / R)}{P(\mathbf{X}, \mathbf{Y} / \bar{R})} \cdot d((\mathbf{X}, \mathbf{Y}))\, \right]$$

$$E_{(\mathbf{X}, \mathbf{Y})}\left[\, \ln \left\{ \frac{P(\mathbf{Y} / \mathbf{X}, R)}{P(\mathbf{Y} / \mathbf{X}, \bar{R})} \cdot \frac{P(\mathbf{X} / R)}{P(\mathbf{X} / \bar{R})} \right\} \cdot d((\mathbf{X}, \mathbf{Y}))\, \right]$$

$$= E_{(\mathbf{X}, \mathbf{Y})}\left[\, \left\{ \ln \frac{P(\mathbf{Y} / \mathbf{X}, R)}{P(\mathbf{Y} / \mathbf{X}, \bar{R})} + \ln \frac{P(\mathbf{X} / R)}{P(\mathbf{X} / \bar{R})} \right\} \cdot d((\mathbf{X}, \mathbf{Y}))\, \right]$$

$$= E_{(\mathbf{X},\ \mathbf{Y})}\left[\ \ln \frac{P(\mathbf{X} / R)}{P(\mathbf{X} / \overline{R})} \cdot d((\mathbf{X},\ \mathbf{Y}))\ \right]$$

$$+ E_{(\mathbf{X},\ \mathbf{Y})}\left[\ \ln \frac{P(\mathbf{Y} / \mathbf{X},\ R)}{P(\mathbf{Y} / \mathbf{X},\ \overline{R})} \cdot d((\mathbf{X},\ \mathbf{Y}))\ \right]. \qquad (3.2.25)$$

Also, from the definition of d,

$$d((\mathbf{X},\ \mathbf{Y})) \equiv d(\mathbf{X}) \equiv d(\mathbf{Y}). \qquad (3.2.26)$$

Therefore, we have

$$S_R((\mathbf{X},\ \mathbf{Y})) = E_{\mathbf{X}}\left[\ \ln \frac{P(\mathbf{X} / R)}{P(\mathbf{X} / \overline{R})} \cdot d(\mathbf{X})\ \right]$$

$$+ E_{\mathbf{X}}\left[\ E_{\mathbf{Y}}\left[\ \ln \frac{P(\mathbf{Y} / \mathbf{X},\ R)}{P(\mathbf{Y} / \mathbf{X},\ \overline{R})} \cdot d((\mathbf{X},\ \mathbf{Y}))\ |\ \mathbf{X}\ \right]\ \right]. \qquad (3.2.27)$$

This leads to (3.2.21)

If $\mathbf{X}$ and $\mathbf{Y}$ are stochastically independent of each other in R and $\overline{R}$, respectively, it is obvious that

$$\rho(\mathbf{Y} \mid \mathbf{X}) \quad \rho(\mathbf{Y}). \qquad (3.2.28)$$

From (3.2.26) and (3.2.28), we have

$$S_R(\mathbf{Y} \mid \mathbf{X}) \quad E_{\mathbf{X}}\left[\ E_{\mathbf{Y}}\left[\ \rho(\mathbf{Y} \mid \mathbf{X}) \cdot d((\mathbf{X},\ \mathbf{Y}))\ |\ \mathbf{X}\ \right]\ \right]$$

$$= E_{(\mathbf{X},\ \mathbf{Y})}\left[\ \rho(\mathbf{Y} \mid \mathbf{X}) \cdot d((\mathbf{X},\ \mathbf{Y}))\ \right]$$

$$= E_{\mathbf{Y}}\left[\ E_{\mathbf{X}}\left[\ \rho(\mathbf{Y} \mid \mathbf{X}) \cdot d((\mathbf{X},\ \mathbf{Y}))\ |\ \mathbf{Y}\ \right]\ \right]$$

$$= E_{\mathbf{Y}}\left[\ E_{\mathbf{X}}\left[\ \rho(\mathbf{Y}) \cdot d(\mathbf{Y})\ \right]\ \right]$$

$$= S_R(\mathbf{Y}), \qquad (3.2.29)$$

from which we obtain (3.2.24). □

We call $S_R(\mathbf{Y} \mid \mathbf{X})$ defined by (3.2.22) *conditional mean information intensity* In Section 3.4.1, we use this quantity to measure the effect of interrelationships between items on the discrimination of sample type R from $\overline{R}$.

In the above, we introduced a measure S_R which evaluates the degree of separation between sample types R and $\overline{R}$, by the use of the quantification ρ.

Now, we consider the quantification of population $\{\mathbf{x}\}$(which is represented by a random variable $\mathbf{X}$ with probability $\tilde{P}(\mathbf{X})$) by the use of ρ. We character-

ize the quantified values of the population $\{\mathbf{x}\}$ by their mean $E_{\tilde{P}(\mathbf{X})}[\rho(\mathbf{X})]$, and by their variance $V_{\tilde{P}(\mathbf{X})}[\rho(\mathbf{X})]$, i.e.,

$$I(R \mid \overline{R} ; \tilde{P}(\mathbf{X})) \triangleq E_{\tilde{P}(\mathbf{X})}[\rho(\mathbf{X})]$$

$$= \sum_{\mathbf{x}} \tilde{P}(\mathbf{X} = \mathbf{x}) \ln \frac{P(\mathbf{X} \quad \mathbf{x} / R)}{P(\mathbf{X} \quad \mathbf{x} / \overline{R})}, \tag{3.2.30}$$

$$V(R \mid \overline{R} ; \tilde{P}(\mathbf{X})) \triangleq V_{\tilde{P}(\mathbf{X})}[\rho(\mathbf{X})]$$

$$= \sum_{\mathbf{x}} \tilde{P}(\mathbf{X} = \mathbf{x}) \cdot \{\ln \frac{P(\mathbf{X} = \mathbf{x} / R)}{P(\mathbf{X} \quad \mathbf{x} / \overline{R})} \quad I(R \mid \overline{R} ; \tilde{P}(\mathbf{X}))\}^2$$

$$= \sum_{\mathbf{x}} \tilde{P}(\mathbf{X} = \mathbf{x}) \cdot \{\ln \frac{P(\mathbf{X} \quad \mathbf{x} / R)}{P(\mathbf{X} \quad \mathbf{x} / \overline{R})}\}^2 \quad I(R \mid \overline{R} ; P(\mathbf{X}))^2. \tag{3.2.31}$$

We call I the *quantified mean* of $\mathbf{X}$ with respect to the sample types R and $\overline{R}$. Also, V is said to be the *quqntified variance* of $\mathbf{X}$ with respect to R and $\overline{R}$. The quantified mean I represents the degree of belongingness of population $\{\mathbf{x}\}$ to R compared to $\overline{R}$.

Particularly, when $\tilde{P}(\mathbf{X}) = P(\mathbf{X} / R)$ or $P(\mathbf{X} / \overline{R})$, then we have

$$I(R \mid \overline{R} ; P(\mathbf{X} / R)) = \sum_{\mathbf{x}} P(\mathbf{X} = \mathbf{x} / R) \cdot \ln \frac{P(\mathbf{X} = \mathbf{x} / R)}{P(\mathbf{X} = \mathbf{x} / \overline{R})}$$

$$= I(P(\mathbf{X} / R) : P(\mathbf{X} / \overline{R}))$$

$$\geq 0, \tag{3.2.32}$$

$$I(R \mid \overline{R} ; P(\mathbf{X} / \overline{R})) = \quad I(P(\mathbf{X} / \overline{R}) : P(\mathbf{X} / R))$$

$$\leq 0, \tag{3.2.33}$$

where $I(P(\mathbf{X} / R) : P(\mathbf{X} / \overline{R}))$ is the so-called Kullback-Leibler information number(cf. Kullback & Leibler[1951], Jeffreys[1946], and Kullback[1952]).

The quantified variance V is used in Section 3.3.2 to analyze the relation of the mean information intensity S_R to the Bayesian discrimination rate.

In the sequel, we consider biochemical data which are commonly used for medical diagnosis. The data are composed of 12 items, each of which represents the content of a biochemical component in human blood(cf. Makino[1971], Takagi, Katai, Iwai et al.[1972], Katai, Iwai et al.[1973 & 1974], Endo & Iwai[1974], and Iwai[1977]). The measurement system is a kind of sequential multiphasic analyzer called SMA 12/60, and the items are as follows:

X_1 : Calcium ion (Ca^{++})

X_2 : Inorganic Phosphorus (Inor. Phos.)
X_3 : Glucose (Glu.)
X_4 : Blood plasma content of Urea Nitrogen (BUN)
X_5 : Uric Acid (Uric Acid)
X_6 : Total Cholesterol (Chol.)
X_7 : Total Protein (T.P.)
X_8 : Albumin (Alb.)
X_9 : Total Bilirubin (T.Bili.)
X_{10}: Alkaline-Phosphotase (Alk.Phos.)
X_{11}: Lactic Dehydrogenase (LDH)
X_{12}: s-Glutamic Oxalacetic Transaminase (SGOT)

Table 3.2.2 is the recording chart, where the grey areas represent the normal regions. The content of each biochemical component X_i is quantized into i_1(less than normal), i_2(normal), and i_3(more than normal) for i = 1, 2, ···, 12 Namely, each item is composed of three categories. The internal diseases are classified into five types(sample types) as follows:

R_1 : Normal
R_2 : Renal failure
R_3 : Liver failure
R_4 : Diabetes
R_5 : the other internal diseases

The data we used were obtained from 532 patients and offered from Kyoto University Hospital. Among them, the numbers of diseases R_1 - R_5 are 90, 73, 165, 69, and 135, respectively. The physiological characteristics of the biochemical components in relation to the diseases are shown by Table 3.3.3(for details, refer to Makino[1971]). We evaluate the effectiveness of each item X_i (i = 1, 2, ···, 12) in relation to the diagnosis of each disease R_j(j = 1, 2, ···, 5) by $S_{R_j}(X_i)$, where R_j $\{R_j, \overline{R}_j\}$ and $\overline{R}_j$ $\bigcup_{k\neq j} R_k$ (cf. Takagi, Katai, Iwai, et al.[1972]). Table 3.2.4 shows the caluculated values of $S_{R_j}(X_i)$(i = 1, 2, ···, 12, j = 1, 2, ···, 5). Comparing the table with Table 3.2.3, we can see that the measure $S_{R_j}(X_i)$ has a good coincidence with the physiological tendencies(cf. Katai, Iwai, et al.[1973]).

In this section, we have introduced a quantification ρ for the category types in a response matrix(table) under the criterion (3.2.3), (3.2.4), and (3.2.5). The quantification $S_R(X_i)$ for item X_i, which is defined by (3.2.5) and is called mean information intensity, can be regarded as a measure of discriminability between types R and $\overline{R}$. The fundamental properties of S_R were

stated by theorem 3.2.1. Moreover, we have introduced a quantification method for a population $\{\mathbf{x}\}$(which is prescribed by its probability law) in terms of the mean I and the variance V of the corresponding quantification ρ. These quantities are, as shown later, closely related to the discrimination rates between the sample types and are available for the item-category aggregation problems in response tables and also for the analyses of certain learning processes.

3.3 Properties of the Measure and Its Relation to Discrimination Rates

In Section 3.2, we have introduced the quantification $S_R(X_i)$(or $S_R(\mathbf{X})$) for each item X_i(or collection of items $\mathbf{X}$), which evaluates the degree of separation between sample types R and $\overline{R}$ with respect to X_i(or $\mathbf{X}$). In other words, $S_R(X_i)$ or $S_R(\mathbf{X})$ can be regarded as the measure of discrimination information contained in X_i or $\mathbf{X}$, respectively. In this section, we consider the properties of the measure S_R, mean information intensity, and its relation to discrimination rates.

In Section 3.3.1, we analyze the properties of the measure with reference to several information measures introduced so far(cf. Kailath[1967]). In Section 3.3.2, we discuss the relation between the measure S_R and the discrimination rate P_c defined by (3.2.30). The relation is derived by the use of the central limit theorem, assuming that the number of items(the number of components in vector $\mathbf{X}$) is sufficiently large and the items are stochastically independent of each other.

3.3.1 Information Theoretical Properties of the Measure

As aforementioned, the measure $S_R(X_i)$ evaluates the amount of information contained in item X_i with respect to the discrimination between sample types R and $\overline{R}$. Hence,it is supposed to have a certain relationship with the information theoretic measures evaluating the discrimination information contained in X_i. In fact, $S_R(X_i)$ is decomposed to three information measures that are wellknown in information theory or statistics, i.e., Shannon's mutual information M, Bhattacharyya distance B, and Kullback-Leibler information number I as follows (cf. Katai & Iwai[1971 & 1972]):

THEOREM 3.3.1 :

$$S_R(X_i) = 2\{B(P(X_i/R), P(X_i/\overline{R})) + M(X_i, R) + I(P(X_i) : P^*(X_i))\}, \tag{3.3.1}$$

where

$$M(X_i, R) \triangleq \sum_{X_i} P(X_i) \ln P(X_i) + \sum_{X_i} P(X_i, R) \ln P(X_i / R)$$

$$+ \sum_{X_i} P(X_i, \bar{R}) \ln P(X_i / \bar{R}) , \qquad (3.3.2)$$

$$B(P(X_i / R) , P(X_i / \bar{R})) \triangleq \ln \gamma_i , \qquad (3.3.3)$$

$$\gamma_i \triangleq \sum_{X_i} \sqrt{P(X_i / R) \cdot P(X_i / \bar{R})} , \qquad (3.3.4)$$

$$I(P(X_i) : P^*(X_i)) \triangleq \sum_{X_i} P(X_i) \ln \frac{P(X_i)}{P^*(X_i)} , \qquad (3.3.5)$$

$$P^*(X_i) \triangleq \frac{\sqrt{P(X_i / R) \cdot P(X_i / \bar{R})}}{\gamma_i} . \qquad (3.3.6)$$

PROOF : From the definition of $S(X_i)$, we have

$$S_R(X_i) \quad \sum_{X_i} \{ P(X_i, R) \ln \frac{P(X_i / R)}{P(X_i / \bar{R})} + P(X_i, \bar{R}) \ln \frac{P(X_i / \bar{R})}{P(X_i / R)} \}$$

$$= \sum_{X_i} \{ P(X_i, R) \ln P(X_i / R) \quad P(X_i, R) \ln P(X_i / \bar{R})$$

$$+ P(X_i, \bar{R}) \ln P(X_i / \bar{R}) \quad P(X_i, \bar{R}) \ln P(X_i / R) \}$$

$$= \sum_{X_i} \{ P(X_i, R) \ln P(X_i / R) \quad (P(X_i) \quad P(X_i, \bar{R})) \ln P(X_i / \bar{R})$$

$$+ P(X_i, \bar{R}) \ln P(X_i / \bar{R}) \quad (P(X_i) \quad P(X_i, R)) \ln P(X_i / R) \}$$

$$= \sum_{X_i} \{ 2 P(X_i, R) \ln P(X_i / R) \quad P(X_i) \ln P(X_i / \bar{R})$$

$$+ \quad 2 P(X_i, \bar{R}) \ln P(X_i / \bar{R}) \quad P(X_i) \ln P(X_i / R) \}$$

$$= 2 \sum_{X_i} \{ - P(X_i) \ln P(X_i) + P(X_i, R) \ln P(X_i / R)$$

$$+ P(X_i, \bar{R}) \ln P(X_i / \bar{R}) \} + \sum_{X_i} \{ 2 P(X_i) \ln P(X_i)$$

$$P(X_i) \ln P(X_i / R) \quad P(X_i) \ln P(X_i / \bar{R}) \}$$

$$= 2M(X_i, R) + 2\sum_{X_i} \{P(X_i) \ln P(X_i) \quad P(X_i) \ln \sqrt{P(X_i / R)}$$

$$- P(X_i) \ln \sqrt{P(X_i / \bar{R})} \}$$

$$= 2 [M(X_i, R) + \sum_{X_i} P(X_i) \{ \ln \frac{P(X_i)}{\sqrt{P(X_i / R)\ P(X_i / \bar{R})}}{\sum_{X_i} \sqrt{P(X_i / R)\ P(X_i / \bar{R})}}$$

$$- \ln \sum_{X_i} \sqrt{P(X_i / R)\ P(X_i / \bar{R})} \}]$$

$$= 2 \{ M(X_i, R) \quad \ln(\sum_{X_i} \sqrt{P(X_i / R)\ P(X_i / \bar{R})})$$

$$+ \sum_{X_i} P(X_i) \ln \frac{P(X_i)}{\sqrt{P(X_i / R)\ P(X_i / \bar{R})}}{\sum_{X_i} \sqrt{P(X_i / R)\ P(X_i / \bar{R})}} \} . \qquad (3.3.7)$$

This leads to (3.3.1). □

These measures are sometimes used for the design of hypothesis testing(cf. Chernoff[1959] and Simons[1967]) and also used for the evaluation of pattern recognition systems(cf. Ho & Agrawara[1968], Caprihan & De Figneiredo[1970], Young[1971], Babu[1972], and Vilmansen[1973]). The interrelationships among the measures were discussed by Kailath(1967) and Toussaint(1972). Also, the geometrical interpretations of the measures and their extensions were given by Rajski(1961), Sibson(1969), Darcóczy(1970), Rathie & Kannappan(1972), and Csiszár(1975).

In the following, the meaning of these three terms in connection with the discrimination between sample types R and $\bar{R}$ is considered.

The first term, Bhattacharyya distance B, does not depend on the *a priori* probabilities P(R) and $P(\bar{R})$ but depends on the conditional probabilities $P(X_i / R)$ and $P(X_i / \bar{R})$. Therefore, the measure B represents the degree of distinguishability between two types R and $\bar{R}$ due to the difference between the probability laws $P(X_i / R)$ and $P(X_i / \bar{R})$. B can be geometrically interpreted as follows(cf.Fig. 3.3.1): Let $\mathbb{P}_1^i$ and $\mathbb{P}_2^i$ be two k_i-dimensional vectors defined for category types $i_1, i_2, \cdots, i_{k_i}$ of item X_i.

$$\mathbf{p}_1^i \triangleq (\sqrt{P(X_i = i_1 / R)}, \sqrt{P(X_i = i_2 / R)}, \cdots, \sqrt{P(X_i \quad i_{k_i} / R)}),$$

$$\mathbf{p}_2^i \triangleq (\sqrt{P(X_i \quad i_1 / \overline{R})}, \sqrt{P(X_i = i_2 / \overline{R})}, \cdots, \sqrt{P(X_i = i_{k_i} / \overline{R})}). \quad (3.3.8)$$

Let these two vectors be at angle θ_i. That is to say, θ_i is a geometrical representation of the relation between the coditional probabilities $P(X_i / R)$ and $P(X_i / \overline{R})$. Then from

$$\| \mathbf{p}_1^i \| = \| \mathbf{p}_2^i \| = 1, \quad (3.3.9)$$

the quantity γ_i defined by (3.3.4) can be represented as

$$\gamma_i = (\mathbf{p}_1^i, \mathbf{p}_2^i) = \cos \theta_i. \quad (3.3.10)$$

Therefore, the relation between B and θ_i is given by

$$B(P(X_i / R), P(X_i / \overline{R})) \qquad \ln \cos \theta_i \geq 0. \quad (3.3.11)$$

Obviously, we have

$$0 \leq \theta_i \leq \frac{\pi}{2}, \text{ i.e., } \gamma_i \geq 0. \quad (3.3.12)$$

Also, it is clear that if $\theta_i = 0$, i.e., no discrimination information with respect to sample types R and $\overline{R}$ can be given by observing item X_i, then $B = 0$. If $\theta_i \quad \pi/2$, i.e., each sample can be completely classified into R or $\overline{R}$ only by the use of information given by X_i, then $B = \infty$.

It is certain that the discrimination information contained in X_i is prescribed not only by the probabilities $P(X_i / R)$ and $P(X_i / \overline{R})$ but also by the *a priori* probabilities $P(R)$ and $P(\overline{R})$. The second and the third terms in (3.3.1) , i.e., Shannon's mutual information M and Kullback-Leibler information number I evaluate the effect of $P(R)$ and $P(\overline{R})$.

The second term M is defined by the difference of entropies as follows(cf. Shannon & Weaver[1949]):

$$M(X_i, R) \triangleq H(R) \quad H(R \mid X_i),$$

$$H(R) \triangleq - P(R) \ln P(R) \quad P(\overline{R}) \ln P(\overline{R}),$$

$$H(R \mid X_i) \triangleq \sum_{X_i} P(X_i, R) \ln P(R / X_i) \quad \sum_{X_i} P(X_i, \overline{R}) \ln P(\overline{R} / X_i). \quad (3.3.13)$$

Let the items $X_1, X_2, \cdots, X_n$ be n signals observed from n distinct communication channels having one input source in common which sends an independently identically distributed(i.i.d.) stochastic sequence of sample type R or $\overline{R}$ according to the *a priori* probabilities P(R) and $P(\overline{R})$(cf. Fig. 3.3.2). The possible number of input sequences of length m is approximated by

$$N = \exp[m \cdot H(\mathcal{R})], \tag{3.3.14}$$

where each sequence has equal probability N^{-1}. The second term $H(\mathcal{R} \mid X_i)$ in $M(X_i, \mathcal{R})$, the so-called equivocation of the i^{th} channel X_i, indicates that the number of possible input sequences subject to a given output sequence form the i^{th} channel is approximately given by

$$N_i = \exp[m \cdot H(\mathcal{R} \mid X_i)], \tag{3.3.15}$$

with equi-probabilities. The above value N_i is bounded by the next inequality(cf. Miyasawa[1971]):

$$H(\mathcal{R} \mid X_i) \leq 2C\sqrt{P(R)\,P(\overline{R})}\,\exp\left[-B(P(X_i / R), P(X_i / \overline{R}))\right], \tag{3.3.16}$$

where

$$C \triangleq \max_{0 \leq x \leq 1} \left[\{ -x \ln x - (1-x) \ln (1-x) \} / \sqrt{x} \right]. \tag{3.3.17}$$

Hence, $H(\mathcal{R} \mid X_i)$ has a certain relationship with the degree of separation B(Bhattacharyya distance). From (3.3.13), the relation among N, N_i, and $M(X_i, \mathcal{R})$ is expressed by

$$\frac{N_i}{N} = \exp[-m \cdot M(X_i, \mathcal{R})]. \tag{3.3.18}$$

Hence, we can say that, while $B(P(X_i / R), P(X_i / \overline{R}))$ evaluates the effectiveness of the i^{th} channel only by the conditional probabilities $P(X_i / R)$ and $P(X_i / \overline{R})$, i.e., without referring to the stochastic properties of the input sequences, mutual information $M(X_i, \mathcal{R})$ evaluates the effectiveness in relation to the actual stochastic properties of the input sequences.

The third term $I(P(X_i) : P^*(X_i))$ can be rewritten as follows:

$$I(P(X_i) : P^*(X_i)) = \sum_{j=1}^{k_i} P(X_i = i_j) \ln L_j^i, \tag{3.3.19}$$

$$L_j^i \triangleq \frac{P(X_i = i_j)}{P^*(X_i = i_j)} \quad \gamma_i \sqrt{P(R)\,P(\overline{R})} \left(\sqrt{T_j^i} + \frac{1}{\sqrt{T_j^i}} \right), \tag{3.3.20}$$

$$T_j^i \triangleq \frac{P(R / X_i = i_j)}{P(\bar{R} / X_i = i_j)}, \quad \text{for } j \quad 1, 2, \cdots, k_i. \tag{3.3.21}$$

T_j^i is the ratio of the *a posteriori* probabilities provided that the sample value of item X_i is equal to i_j. Its logarithmic value can be considered as the amount of information given by the sample value i_j. The measure $I(P(X_i) : P^*(X_i))$ has the minimum value 0 when $L_j^i = 1$ for all $j = 1, 2, \cdots, k_i$, i.e., when the ratio T_j^i has values

$$T_j^i \quad \bar{T}^i \text{ or } \frac{1}{\bar{T}^i} \quad \text{for all } j = 1, 2, \cdots, k_i, \tag{3.3.22}$$

where $\bar{T}^i$ is defined as

$$\bar{T}^i \triangleq \frac{1 + \sqrt{1 \quad 4\gamma_i^2 p(R) P(\bar{R})}}{2\gamma_i \sqrt{P(R) P(\bar{R})}}. \tag{3.3.23}$$

Therefore, the third component I of $S_R(X_i)$ increases along with the increase of variation among the amount of information $|\ln T_1^i|, |\ln T_2^i|, \cdots, |\ln T_{k_i}^i|$. In the Bayesian discrimination, the decision of sample types is made according to whether the value $\ln T_j^i$ is positive or negative. Namely, compared with the second term, the mutual information M which represents the mean efficiency of the information transmission of the i^{th} channel X_i, the third term I represents the effectiveness of the manner of appearence of sample values of X_i.

In this section, a new information measure, mean imformation intensity $S_R(X_i)$, is defined and shown to be composed of three information theoretic measures, Bhattacharyya distance, mutual information, and Kullback-Leibler information number. We also investigated the meaning of these three components in relation to the discrimination between sample types R and $\bar{R}$. It was shown that each of the components only represents each aspect of discrimination information contained in X_i. Hence the validity of the measure $S_R(X_i)$ should also be examined in connection with the discrimination rates.

3.3.2 Approximation of Discrimination Rates of Likelihood Ratio Tests and Cosideration on Their Relation with the Measure

As aforementioned, the discrimination rates of likelihood ratio tests can be represented as (3.2.20) by the use of threshold function s(t;k). In this section, we confine ourselves to the case of the Bayesian discrimination. In

this case, the critical value k is

$$k = \ln \frac{P(R)}{P(\bar{R})} . \tag{3.3.24}$$

Let us transform the quantification $\rho(\mathbf{X})$ into $\bar{\rho}(\mathbf{X})$ as follows(cf. Katai & Iwai [1971]):

$$\begin{aligned} \bar{\rho}(\mathbf{X}) &\triangleq \ln \frac{P(R / \mathbf{X})}{P(\bar{R} / \mathbf{X})} \\ &= \ln \frac{P(\mathbf{X}, R)}{P(\mathbf{X}, \bar{R})} \\ &= \ln \frac{P(\mathbf{X} / R) \cdot P(R)}{P(\mathbf{X} / \bar{R}) \cdot P(\bar{R})} \\ &= \rho(\mathbf{X}) + k. \end{aligned} \tag{3.3.25}$$

Then the critical value corresponding to $\bar{\rho}(\mathbf{X})$ is equal to 0 and the discrimination rate P_c is given by the next formula:

$$P_c = \frac{1}{2} + \frac{1}{2} E_{\mathbf{X}}[\, s(\bar{\rho}(\mathbf{X});0) \cdot d(\mathbf{X})]. \tag{3.3.26}$$

As the first approximation of P_c, we first investigate the quantity $E_{\mathbf{X}}[\bar{\rho}(\mathbf{X}) \cdot d(\mathbf{X})]$ that is given by approximating the nonlinear threshold function $s(t;0)$ by t(cf. Fig. 3.3.3). From (3.3.25), we have

$$\begin{aligned} E_{\mathbf{X}}[\bar{\rho}(\mathbf{X}) \cdot d(\mathbf{X})] \quad & E_{\mathbf{X}}[(\rho(\mathbf{X}) + k) \cdot d(\mathbf{X})] \\ &= E_{\mathbf{X}}[\rho(\mathbf{X}) \cdot d(\mathbf{X})] + k \cdot E_{\mathbf{X}}[d(\mathbf{X})] \\ & S_R(\mathbf{X}) + J(P(R), P(\bar{R})), \end{aligned} \tag{3.3.27}$$

where

$$\begin{aligned} J(P(R), P(\bar{R})) &\triangleq k \cdot E_{\mathbf{X}}[d(\mathbf{X})] \\ & (P(R) \quad P(\bar{R})) \ln \frac{P(R)}{P(\bar{R})} \\ &= P(R) \ln \frac{P(R)}{P(\bar{R})} + P(\bar{R}) \ln \frac{P(\bar{R})}{P(R)}. \end{aligned} \tag{3.3.28}$$

The term J is the Hájek's divergence between the *a priori* probabilities P(R) and $P(\bar{R})$(cf. Katai, Iwai, et al.[1975c]). It evaluates the difference between P(R)

and $P(\overline{R})$, i.e., the amount of *a priori* information due to the difference between $P(R)$ and $P(\overline{R})$(cf. Fig. 3.3.4). The quantity $E_{\mathbf{X}}[\overline{\rho}(\mathbf{X})\cdot d(\mathbf{X})]$ is equal to the measure $S_R(\mathbf{X})$ plus *a priori* information $J(P(R), P(\overline{R}))$.

As shown in the previous section, the measure $S_R(\mathbf{X})$ can be decomposed into the three information theoretic measures, and there have been many studies concerning the relationships between the discrimination rates(error probabilities) and the values of the above measures(cf. Hellman & Raviv[1970], Toussaint [1971], Babu[1972], and Vilmansen[1973]). However, these studies only show the relationships between the two quantities as inequalities. Hence, in the following , we derive an approximation of the Bayesian discrimination rate (3.2.20) by the use of the central limit theorem under the assumption that the items X_1, X_2, $\cdots$, and X_n are stochastically independent in each of the sample types R and $\overline{R}$.

Let P_R and $P_{\overline{R}}$ be the discrimination rates of the first kind with respect to R and $\overline{R}$, respectively, i.e.,

$$P_R = \text{Prob.}(\ \mathbf{X} \Rightarrow \text{decision } R \,/\, \mathbf{X} \in R),$$

$$P_{\overline{R}} = \text{Prob.}(\ \mathbf{X} \Rightarrow \text{decision } \overline{R} \,/\, \mathbf{X} \in \overline{R}). \tag{3.3.29}$$

Bayesian discrimination rate P_c is expressed as a weighted average of the above two quantities, i.e.,

$$P_c = P(R)\cdot P_R + P(\overline{R})\cdot P_{\overline{R}}\ . \tag{3.3.30}$$

From the definition of P_R and $P_{\overline{R}}$, we have

$$P_R = \text{Prob.}(\ \overline{\rho}(\mathbf{X}) > 0 \,/\, \mathbf{X} \in R),$$

$$P_{\overline{R}} = \text{Prob.}(\ \overline{\rho}(\mathbf{X}) < 0 \,/\, \mathbf{X} \in \overline{R}). \tag{3.3.31}$$

Also, from (3.3.25) and the stochastic independence of X_1, X_2, $\cdots$, and X_n, $\overline{\rho}(\mathbf{X})$ is rewritten as

$$\overline{\rho}(\mathbf{X}) = \ln \frac{P(R)}{P(\overline{R})} + \sum_{i=1}^{n} \rho(X_i). \tag{3.3.32}$$

The random variables $\rho(X_1)$, $\rho(X_2)$, $\cdots$, and $\rho(X_n)$ are stochastically independent in each of the sample types, from the assumption. Therefore, if the number of items n is sufficiently large, the quantity $\overline{\rho}(\mathbf{X})$ has approximately a normal distribution in each of the types as follows.

In type R:

$$\bar{\rho}(\mathbf{X}) \sim N\left(\ln \frac{P(R)}{P(\bar{R})} + I(R \mid \bar{R} \; ; \; P(\mathbf{X} / R)),\; V(R \mid \bar{R} \; ; \; P(\mathbf{X} / R)) \right), \tag{3.3.33}$$

where

$$I(R \mid \bar{R} \; ; \; P(\mathbf{X} / R)) = \sum_{i=1}^{n} I(R \mid \bar{R} \; ; \; P(X_i / R)), \tag{3.3.34}$$

$$V(R \mid \bar{R} \; ; \; P(\mathbf{X} / R)) \quad \sum_{i=1}^{n} V(R \mid \bar{R} \; ; \; P(X_i / R)). \tag{3.3.35}$$

In type $\bar{R}$:

$$\bar{\rho}(\mathbf{X}) \sim N\left(-\ln \frac{P(\bar{R})}{P(R)} \quad I(\bar{R} \mid R \; ; \; P(\mathbf{X} / \bar{R})), V(\bar{R} \mid R \; ; \; P(\mathbf{X} / \bar{R})) \right), \tag{3.3.36}$$

where

$$I(\bar{R} \mid R \; ; \; P(\mathbf{X} / \bar{R})) \quad \sum_{i=1}^{n} I(\bar{R} \mid R \; ; \; P(X_i / \bar{R})), \tag{3.3.37}$$

$$V(\bar{R} \mid R \; ; \; P(\mathbf{X} / \bar{R})) \quad \sum_{i=1}^{n} V(\bar{R} \mid R \; ; \; P(X_i / \bar{R})). \tag{3.3.38}$$

In the above, the quantities I and V are introduced in Section 3.2(cf. (3.2.30) and (3.2.31)), and the additivity relations (3.3.34), (3.3.35), (3.3.37), and (3.3.38) are due to the assumption of stochastic independence. Therefore, the discrimination rate of the first kind can be approximated as

$$P_R \sim \frac{1}{2} + \frac{1}{2} \Phi\left(\frac{\sum_{i=1}^{n} I(R \mid \bar{R} \; ; \; P(X_i / R)) + \ln P(R) \quad \ln P(\bar{R})}{\left(2 \sum_{i=1}^{n} V(R \mid \bar{R} \; ; \; P(X_i / R)) \right)^{1/2}} \right),$$

$$P_{\bar{R}} \simeq \frac{1}{2} + \frac{1}{2} \Phi\left(\frac{\sum_{i=1}^{n} I(\bar{R} \mid R \; ; \; P(X_i / \bar{R})) + \ln P(\bar{R}) \quad \ln P(R)}{\left(2 \sum_{i=1}^{n} V(\bar{R} \mid R \; ; \; P(X_i / \bar{R})) \right)^{1/2}} \right), \tag{3.3.39}$$

where Φ is the so-called error function defined by

$$\Phi(x) \triangleq \frac{1}{\sqrt{\pi}} \int_{-x}^{x} \exp(-t^2)\, dt. \tag{3.3.40}$$

From the definition of $S_R(X_i)$ and (3.2.30), we have

$$S_R(X_i) = E_{X_i}[\rho(X_i) \cdot d(X_i)]$$

$$= P(R)\, E_{X_i}[\rho(X_i) \mid R] + P(\bar{R})\, E_{X_i}[\rho(X_i) \mid \bar{R}]$$

$$P(R)\, I(R \mid \overline{R}\ ;\ P(X_i / R)) + P(\overline{R})\, I(\overline{R} \mid R\ ;\ P(X_i / \overline{R})). \qquad (3.3.41)$$

The quantity $S_{\mathcal{R}}$ is concerned only with the quantified means $I(R \mid \overline{R}\ ;\ P(X_i / R))$ and $I(\overline{R} \mid R\ ;\ P(X_i / \overline{R}))$, and not with the quantified variances $V(R \mid \overline{R}\ ;\ P(X_i / R))$ and $V(\overline{R} \mid R\ ;\ P(X_i / \overline{R}))$.

As a simple illustration of the above discussion, we consider the following case: Each item X_i of $\mathbf{X}$ is either i_0 or i_1 which originates from the ideal patterns $(1_0, 2_0, \cdots, i_0, \cdots, n_0)$ (for type R) or $(1_1, 2_1, \cdots, i_1, \cdots, n_1)$ (for type $\overline{R}$) such that

$$P(X_i \quad i_1 / R) = p,\ P(X_i \quad i_0 / R) = 1 \quad p,$$

$$P(X_i \quad i_1 / \overline{R}) = 1 \quad p,\ P(X_i = i_0 / \overline{R}) = p. \qquad (3.3.42)$$

Namely, noise content ratio has the constant value p in both types R and $\overline{R}$ and also in all the items $X_1, X_2, \cdots$, and X_n. These probability structures are illustrated in Table 3.3.1.

Supposing that

$$p < \frac{1}{2}, \qquad (3.3.43)$$

we calculate the values of $S_{\mathcal{R}}(X_i)$, $M(X_i, \mathcal{R})$, $B(P(X_i / R), P(X_i / \overline{R}))$, $I(P(X_i) : P^*(x_i))$, $I(R \mid \overline{R}\ ;\ P(X_i / R))$, $V(R \mid \overline{R}\ ;\ P(X_i / R))$, P_c etc.

In this case, we have

$$I(R \mid \overline{R}\ ;\ P(X_i / R)) = p \ln\frac{p}{1 - p} + (1 \quad p) \ln\frac{1 \quad p}{p}$$

$$= (1 - 2p) \ln\frac{1 \quad p}{p}$$

$$I(\overline{R} \mid R\ ;\ P(X_i / \overline{R}))$$

$$\triangleq I_{\mathcal{R}}(p). \qquad (3.3.44)$$

It follows that

$$S_{\mathcal{R}}(X_i) = I_{\mathcal{R}}(p) = (1 - 2p) \ln\frac{1 \quad p}{p}. \qquad (3.3.45)$$

From (3.3.2), Shannon's mutual information M is given by the next equation, where r and 1 r are the *a priori* probabilities of R and $\overline{R}$, respectively.

$$M(X_i, \mathcal{R}) = \{(2p \quad 1) r + (1 \quad r)\} \ln\{(2p \quad 1) r + (1 \quad r)\}$$

$$\{(1 \quad 2p)r + p\} \ln\{(1 \quad 2p)r + p\}$$

$$+ p \ln p + (1 \quad p) \ln (1 \quad p)$$

$$\triangleq M(r, p) . \tag{3.3.46}$$

The quantity $M(r, p)$ has its maximum value when $r = 1/2$ as follws.

$$M(\frac{1}{2}, p) = \max_{0 \le r \le 1} M(r, p)$$

$$= p \ln p + (1 \quad p) \ln (1 \quad p) + \ln 2$$

$$= M(x_i, R) + I(P(X_i) : P^*(X_i)). \tag{3.3.47}$$

Bhattacharyya distance $B(P(X_i / R) , P(X_i / \overline{R}))$ is

$$B(P(X_i / R) , P(X_i / \overline{R})) = \ln (\sqrt{p}\sqrt{1 \quad p} + \sqrt{1 \quad p}\sqrt{p})$$

$$= \ln 2 \quad \frac{1}{2} \ln\{p(1 \quad p)\}$$

$$\triangleq B(p) \tag{3.3.48}$$

From theorem 3.3.1, $S_R(p)$ can be expressed by

$$S_R(p) \quad 2B(p) + 2M(\frac{1}{2}, p). \tag{3.3.49}$$

Fig. 3.3.5 shows the values of $S_R(p)$, $2B(p)$, and $2M(\frac{1}{2}, p)$ versus p. Roughly speaking, $S_R(p)$ and $B(p)$ have linear relationships with $\ln p$. Also, we obtain that

$$\text{if} \quad p \to \frac{1}{2}, \quad \text{then} \quad \frac{S_R(p)}{M(\frac{1}{2}, p)} \to 4, \quad \text{and}$$

$$\text{if} \quad p \to 0, \quad \text{then} \quad \frac{S_R(p)}{M(\frac{1}{2}, p)} \to \infty. \tag{3.3.50}$$

Next, we investigate the relation of the discrimination rate P_c with the probability p and the number of items n. From (3.2.31),

$$V(R \mid \overline{R} \; ; \; P(X_i / R)) = 4p(1 \quad p)\{\ln \frac{1 \quad p}{p}\}^2$$

$$= V(\overline{R} \mid R \; ; \; P(X_i / \overline{R}))$$

$$\triangleq V_R(p). \tag{3.3.51}$$

Assume that

$$P(R)\ (\ =\ r\ )\quad \frac{1}{2}\ . \tag{3.3.52}$$

Then, from (3.3.39), discrimination rate P_c is given as

$$\begin{aligned} P_c \quad & P_R \\ & P_{\overline{R}} \\ &= \frac{1}{2} + \frac{1}{2}\,\Phi(\sqrt{n}\cdot\tau(p)) \\ &\triangleq P_c(n,\ p)\ . \end{aligned} \tag{3.3.53}$$

In the above equation, $\tau(p)$, which represents the signal to noise ratio of the random variable $\rho(X_i)$, is defined as follows:

$$\begin{aligned} \tau(p) &\triangleq \frac{S_R(p)}{\sqrt{V_R(p)}} \\ &= \frac{1 \quad 2p}{2\sqrt{p\,(1\ -\ p)}}\ . \end{aligned} \tag{3.3.54}$$

The relation between $\tau(p)$ and p is illustrated in Fig. 3.3.6. Fig. 3.3.7 shows the relation of $P_c(n,\ p)$ to n and p. From the figure, we can estimate the necessary number of items to assure arbitrarily given discrimination rates.

In order to evaluate the effect of the i^{th} item X_i on the Bayesian discrimination, we calculate the quantity $\Delta_i P_R$(or $\Delta_i P_{\overline{R}}$), the decrease of the Bayesian discrimination rate P_R(or $P_{\overline{R}}$) due to the deletion of X_i from $\mathbf{X}$, i.e.,

$$\Delta_i P_R \triangleq P_R \quad P_R{}^i\ , \tag{3.3.55}$$

where $P_R{}^i$ is the discrimination rate of the first kind based on the information given by $\mathbf{X}' \triangleq (X_1,\ X_2,\ \cdots,\ X_{i-1},\ X_{i+1},\ \cdots,\ X_n)$. According to the next formula

$$\frac{\beta}{\sqrt{\alpha}} - \frac{\beta \quad \Delta\beta}{\sqrt{\alpha} \quad \Delta\alpha} \simeq \frac{1}{\sqrt{\alpha}}\,(\Delta\beta \quad \frac{\beta}{2\alpha}\,\Delta\alpha) \tag{3.3.56}$$

and also to the following equality

$$\frac{d\Phi(x)}{dx} = \frac{2}{\sqrt{\pi}}\exp(-x^2), \tag{3.3.57}$$

we obtain

$$\Phi\left(\frac{\beta}{\sqrt{\alpha}}\right) \quad \Phi\left(\frac{\beta \quad \Delta\beta}{\sqrt{\alpha \quad \Delta\alpha}}\right) \simeq \frac{2}{\sqrt{\alpha\pi}} \exp\left(-\frac{\beta^2}{\alpha}\right) \cdot \left(\Delta\beta \quad \frac{1}{2}\frac{\beta}{\alpha}\Delta\alpha\right). \tag{3.3.58}$$

Therefore, from (3.3.39), the above quantity $\Delta_i P_R$ can be approximated as

$$\Delta_i P_R \simeq F_R \cdot G_R(X_i), \tag{3.3.59}$$

where

$$F_R \triangleq \frac{1}{\sqrt{2\pi n \bar{V}_R}} \exp\left\{-\frac{1}{2}\left(\frac{n\bar{I}_R + \ln P(R) \quad \ln P(\bar{R})}{\sqrt{n\bar{V}_R}}\right)^2\right\}, \tag{3.3.60}$$

$$G_R(X_i) \triangleq I(R \mid \bar{R} \; ; \; P(X_i / R)) \quad \frac{1}{2}\frac{\bar{I}_R}{\bar{V}_R} V(R \mid \bar{R} \; ; \; P(X_i / R)), \tag{3.3.61}$$

$$\bar{I}_R \triangleq \frac{1}{n} \sum_{j=1}^{n} I(R \mid \bar{R} \; ; \; P(X_j / R)), \tag{3.3.62}$$

$$\bar{V}_R \triangleq \frac{1}{n} \sum_{j=1}^{n} V(R \mid \bar{R} \; ; \; P(X_j / R)). \tag{3.3.63}$$

The quatities $\bar{I}_R$ and $\bar{V}_R$ are the mean values of quantified means $I(R \mid \bar{R} \; ; \; P(X_j / R))(j = 1, 2, \cdots, n)$ and quantified variances $V(R \mid \bar{R} \; ; \; P(X_j / R))$ $(j \quad 1, 2, \cdots, n)$, respectively. Thus the ratio $\bar{I}_R / \bar{V}_R$ represents the accuracy or the reliability of the information gained from the vector $\mathbf{X}$. From (3.3.59), the quantity $G_R(X_i)$ represents the relative effect of item X_i with respect to the discrimination rate P_R. From (3.3.61), the quatified mean $I(R \mid \bar{R} \; ; \; P(X_i / R))$ has a positive effect on $G_R(X_i)$, while the quantified variance $V(R \mid \bar{R} \; ; \; P(X_i / R))$ has a negative effect whose amount is dependent on the accuracy $\bar{I}_R / \bar{V}_R$ of the information given by $\mathbf{X}$.

If we measure the relative effectiveness of the discrimination information contained in X_i by the weighted average of $G_R(X_i)$ and $G_{\bar{R}}(X_i)$ with *a priori* probabilities $P(R)$ and $P(\bar{R})$ as follows,

$$G_R(X_i) = P(R) \cdot G_R(X_i) + P(\bar{R}) \cdot G_{\bar{R}}(X_i), \tag{3.3.64}$$

then, from (3.3.41) and (3.3.61), the above measure can be rewritten as

$$G_R(X_i) = S_R(X_i) \quad K_R(X_i), \tag{3.3.65}$$

where

$$K_R(x_i) \triangleq \frac{1}{2}\{ P(R)\frac{\overline{I}_R}{\overline{V}_R}V(R \mid \overline{R} \ ; \ P(X_i / R)) + P(\overline{R})\frac{\overline{I}_{\overline{R}}}{\overline{V}_{\overline{R}}}V(\overline{R} \mid R \ ; \ P(X_i / \overline{R}))\}. \tag{3.3.66}$$

Thus the mean information intensity $S_R(X_i)$ positively contributes to the relative effectiveness $G_R(X_i)$. The negative contribution $K_R(X_i)$ depends on the quatified variances $V(R \mid \overline{R} \ ; \ P(X_i / R))$ and $V(\overline{R} \mid R \ ; \ P(X_i / \overline{R}))$ and also on the accuracies $\overline{I}_R / \overline{V}_R$ and $\overline{I}_{\overline{R}} / \overline{V}_{\overline{R}}$ of the information given by **X**. If the accuracies are high, then the negative contribution $K_R(X_i)$ is large. Hence, it can be regarded as a measure of the information contained in **X** which cancels the information (whose amount is evaluated by $S_R(X_i)$) contained in X_i.

As stated above, there is no direct relation between the relative effctiveness $G_R(X_i)$ and the information measure $S_R(X_i)$. However, in the following particular cases, $G_R(X_i)$ has an approximate linear relation with $S_R(X_i)$(cf. Katai & Iwai[1972]).

THEOREM 3.3.2 : In each of the following two cases a and b, there exists an approximate linear relationship between $G_R(X_i)$ and $S_R(X_i)$, i.e.,

$$G_R(X_i) \simeq \frac{1}{2} S_R(X_i), \tag{3.3.67}$$

and $S_R(X_i)$ can be considered as a useful measure of the discrimination information contained in X_i.

a) All the items X_1, X_2, $\cdots$, and X_n have approximately the same probability structures, i.e.,

$$P(X_i / R) \simeq P(X_j / R),$$

$$P(X_i / \overline{R}) \simeq P(X_j / \overline{R}), \quad \text{for all } i, j = 1, 2, \cdots, n. \tag{3.3.68}$$

b) Sample types R and $\overline{R}$ have approximately the same probability structures, i.e.,

$$P(X_j / R) \simeq P(X_j / \overline{R}) \quad \text{for all } j = 1, 2, \cdots, n, \tag{3.3.69}$$

more precisely,

$$|P(X_j / R) \quad P(X_j / \overline{R})| \ll P(X_j / R) \text{ for } j \quad 1, 2, \cdots, n. \tag{3.3.70}$$

PROOF : In case a, we have

$$I(R \mid \overline{R} \ ; \ P(X_j / R)) \simeq I(R \mid \overline{R} \ ; \ P(X_i / R)),$$

$$I(\bar{R} \mid R\ ;\ P(X_j / \bar{R})) \simeq I(\bar{R} \mid R\ ;\ P(X_i / \bar{R})),$$

$$V(R \mid \bar{R}\ ;\ P(X_j / R)) \simeq V(R \mid \bar{R}\ ;\ P(X_i / R)),$$

$$V(\bar{R} \mid R\ ;\ P(X_j / \bar{R})) \simeq V(\bar{R} \mid R\ ;\ P(X_i / \bar{R})), \text{for } j \quad 1, 2, \cdots, n. \quad (3.3.71)$$

Also, from (3.3.66), we obtain

$$K_R(X_i) \sim \frac{1}{2}\{P(R) \frac{n \cdot I(R \mid \bar{R}\ ;\ P(X_i / R))}{n \cdot V(R \mid \bar{R}\ ;\ P(X_i / R))} V(R \mid \bar{R}\ ;\ P(X_i / R))$$

$$+ P(\bar{R}) \frac{n \cdot I(\bar{R} \mid R\ ;\ P(X_i / \bar{R}))}{n \cdot V(\bar{R} \mid R\ ;\ P(X_i / \bar{R}))} V(\bar{R} \mid R\ ;\ P(X_i / \bar{R}))\}$$

$$= \frac{1}{2}\{P(R) \cdot I(R \mid \bar{R}\ ;\ P(X_i / R)) + P(\bar{R}) \cdot I(\bar{R} \mid R\ ;\ P(X_i / \bar{R}))\}$$

$$= \frac{1}{2} S_R(X_i). \quad (3.3.72)$$

For case b. Let $\alpha(X_j)$ denote

$$\alpha(X_j) \triangleq \frac{P(X_j / \bar{R}) \quad P(X_j / R)}{P(X_j / R)}. \quad (3.3.73)$$

Then, from condition (3.3.70),

$$\alpha(X_j) \ll 1 \quad \text{for } j = 1, 2, \cdots, n. \quad (3.3.74)$$

By the use of Taylor expansion, we have

$$\ln \frac{P(X_j / R)}{P(X_j / \bar{R})} = \alpha(X_j) + \frac{1}{2}\alpha(X_j)^2 + O(\alpha^3)$$

$$\frac{P(X_j / R) \quad P(X_j / \bar{R})}{P(X_j / R)} + \frac{1}{2} \frac{(P(X_j / R) \quad P(X_j / \bar{R}))^2}{P(X_j / R)^2} + O(\alpha^3). \quad (3.3.75)$$

From (3.2.30), we obtain

$$I(R \mid \bar{R}\ ;\ P(X_j / R)) \quad \sum_{X_j} P(X_j / R)\{ \frac{P(X_j / R) - P(X_j / \bar{R})}{P(X_j / R)}$$

$$+ \frac{1}{2} \frac{(P(X_j / R) \quad P(X_j / \bar{R}))^2}{P(X_j / R)^2} \} + O(\alpha^3)$$

$$= \sum_{X_j} \{P(X_j / R) \quad P(X_j / \bar{R})\} + \frac{1}{2} \sum_{X_j} \frac{(P(X_j / R) \quad P(X_j / \bar{R}))^2}{P(X_j / R)} + O(\alpha^3). \quad (3.3.76)$$

Noting that

$$\sum_{x_j} P(X_j / R) = \sum_{X_j} P(X_j / \bar{R}) - 1 \quad (3.3.77)$$

and also disregarding the terms of $O(\alpha^3)$, we obtain

$$I(R \mid \bar{R} \; ; \; P(X_j / R)) \simeq \frac{1}{2} \sum_{X_j} \frac{(P(X_j / R) \quad P(X_j / \bar{R}))^2}{P(X_j / R)} \quad (= O(\alpha^2)) . \quad (3.3.78)$$

On the other hand, from (3.2.31) and (3.3.75),

$$V(R \mid \bar{R} \; ; \; P(X_j / R)) \quad \sum_{X_j} P(X_j / R) \{ \frac{P(X_j / R) - P(X_j / \bar{R})}{P(X_j / R)} + O(\alpha^2) \}^2$$

$$I(R \mid \bar{R} \; ; \; P(X_j / R))^2$$

$$\sum_{x_j} \frac{(P(X_j / R) \quad P(X_j / \bar{R}))^2}{P(X_j / R)} + O(\alpha^3). \quad (3.3.79)$$

Therefore, we obtain

$$V(R \mid \bar{R} \; ; \; P(X_j / R)) \simeq 2 I(R \mid \bar{R} \; ; \; P(X_j / R)),$$

$$V(\bar{R} \mid R \; ; \; P(X_j / \bar{R})) \simeq 2 I(\bar{R} \mid R \; ; \; P(X_j / \bar{R})), \quad \text{for } j = 1, 2, \cdots, n. \quad (3.3.80)$$

These lead to

$$K_R(X_i) \simeq \frac{1}{2} \{P(R) \cdot I(R \mid \bar{R} \; ; \; P(X_i / R)) + P(\bar{R}) \cdot I(\bar{R} \mid R \; ; \; P(X_i / \bar{R}))\}$$

$$= \frac{1}{2} S_R(X_i). \; \square$$

Experimental considerations by computer simulations

In the above, we introduced an information theoretic measure $S_R(X_i)$ and showed that it is composed of three components, Bhattacharyya distance B, Kullback-Leibler information number I, and Shannon's mutual information M. We

also derived the relation between the measure S_R and G_R, the measure of relative effectiveness based on the Bayesian discrimination rates, under certain conditions and approximations.

In the sequel, we examine the validity of the above approximations and the properties of the three components by computer simulations(cf. Katai & Iwai [1972]).

As shown in Fig. 3.3.8, the example model is composed of 4 × 4 = 16 items, i.e.,

$$\mathbf{X} = (X_{11}, X_{12}, \cdots, X_{41}, \cdots, X_{44}). \tag{3.3.81}$$

By quantizing a random variable Z_{ij} with four points -D, -C, C, and D, each item X_{ij} takes one of five possible categories(category types) ij_1, ij_2, $\cdots$, ij_5. The random variable Z_{ij} has normal distribution N(0, 1) in the type R and has $N(\mu_{ij}, \sigma_{ij})$ in the type $\overline{R}$. The means μ_{ij}'s and the standard deviations σ_{ij}'s are set as follows:

$$\mu_{ij} = \frac{RAM}{3}(i \quad 1), \quad \sigma_{ij} = \frac{RAV}{3}(j - 1) + 1, \text{ for } i,j \quad 1,2,3,4. \tag{3.3.82}$$

Hence, all the items have different probability structures and the degree of separation of R from $\overline{R}$ can be set arbitrarily by changing the values of RAM and RAV. We used 3,000 samples(as test inputs) to estimate the discrimination rates P_R, $P_{\overline{R}}$, and P_c, and their amount of decrease $\Delta_{ij}P_R$, $\Delta_{ij}P_{\overline{R}}$, and $\Delta_{ij}P_c$. The parameters were set as $P(R) = P(\overline{R}) = 1/2$, C = 0.3, and D 0.9.

The results of the computer simulations are as follows:

1) On the validity of the approximate formula (3.3.39).

Equation (3.3.39) is derived from the central limit theorem under the condition that $\{X_{ij}\}$ are stochastically independent and that the number of the items n is sufficiently large. However, when the number n is large, the decrease of the discrimination rates $\Delta_{ij}P_R$ and $\Delta_{ij}P_{\overline{R}}$ become too small to assure the statistical reliability of their experimental values. Therefore, the number n cannot be set so large. So, in this case, we set n = 16. Nevertheless, the theoretical values of the discrimination rates by the above approximation have quite a good coincidence with the experimental values(cf. Table 3.3.2). Thus, formula (3.3.39) is considered to be a good approximation of the true discrimination rate.

2) On the precision of approximation (3.3.59).

We calculated the theoretical and the experimental values of $\Delta_{ij}P_c$ $P(R)\cdot\Delta_{ij}P_R$ + $P(\overline{R})\cdot\Delta_{ij}P_{\overline{R}}$ $\frac{1}{2}\{\Delta_{ij}P_R + \Delta_{ij}P_{\overline{R}}\}$ for the representative items X_{41}, X_{14}, and X_{44}(

cf. Fig. 3.3.9), where the corresponding random variables Z_{41}, Z_{14}, and Z_{44} have probability distributions(in type $\bar{R}$) as follows:

$$Z_{41} \sim N(RAM, 1),$$

$$Z_{14} \sim N(0, RAV),$$

$$Z_{44} \sim N(RAM, RAV). \qquad (3.3.83)$$

In the figure, there exist considerable differences between the theoretical and the experimental values of $\Delta_{ij}P_c$. However, the essential trend of the experimental values coincide with that of the theoretical values. The difference between them is considered to be due to a statistical fluctuation because of the number of the samples being insufficient.

3) On the relation (3.3.67) between the relative effectiveness $G_R(X_i)$ and the mean information intensity $S_R(X_i)$.

The validity of the approximate formula (3.3.67) under the condition b in theorem 3.3.2 is examined providing that RAM and RAV are small. From Fig. 3.3.10, the approximate formula is valid even when RAM and RAV are considerably large(when the condition (3.3.70) does not hold).

4) On the characteristics of the mean information intensity S_R and its three components B, M, and I.

Fig. 3.3.11 shows the values of S_R and the three components of S_R (Bhattacharyya distance B, Shannon's mutual information M, and Kullback-Leibler information number I) for various values of the mean μ_{ij} and the standard deviation σ_{ij}. In the figure, B and M have similar values and are more sensitive to the difference of the mean values ($\mu \triangleq \mu_{ij}$ 0) than the ratio of the standard deviations ($\sigma \triangleq \sigma_{ij}$ / 1). In this case, S_R is approximately four times B or M. The values of the Kullback-Leibler information number I are quite small compared with B or M except for the item X_{41}. The item is peculiar in that the difference of means μ is maximum and the ratio of standard deviation σ is minimum. Hence, in the item, the values of $| \ln T_j^{41} |$ $| \ln(P(R / X_{41} \quad 41_j) / P(\bar{R} / X_{41} - 41_j)) |$ have the most inhomogeneity with respect to j, making the value of I the largest.

In this section, we have discussed the relationship between the mean information intensity S_R and the Bayesian discrimination rates. In general, as shown by (3.3.39), the Bayesian discrimination rates are prescribed by the quantified means and quantified variances introduced in Section 3.2. The relationship between the mean information intensity S_R and the quantified means was shown by (3.3.41). In order to examine the above relationships, we intro-

duced a measure G_R which evaluates the effect of each item on the discrimination rate. According to theorem 3.2.2, the mean information intensity S_R has approximately a linear relationship with G_R under certain conditions. Taking the useful properties indicated by theorem 3.2.2 into account, the measure S_R is considered to be highly applicable to practical use. Also, the computer simulations assure the validity of the approximately linear relationship.

3.4 Considerations on Item-Category Aggregation Problems based on the Measure and the Quantification

In the preceding sections , we introduced the quantification $\rho(x_i)$ and the measure S_R of information content in each item X_i and also examined the properties of the measure under the assumption of stochastic independence among the items in response matrices. It is obvious that the discrimination rates are closely related to the item-category classification of response matrices. In this section, we investigate the efficient way of item-cagegory aggregation based on the measure, mean information intensity S_R

In Section 3.4.1, the problem of item aggregation is considered introducing a measure of correlation between items with respect to discrimination by response matrices. In Section 3.4.2, the problem of category aggregation is treated in a more general setting of the problem, introducing a measure of information loss.

3.4.1 Introduction of a Measure of Correlation between Items and Investigation of the Efficient Way of Item Aggregation

In this section, we introduce a measure of correlation between items for the determination of effective item aggregation in the case where the items are not stochastically independent. Lewis(1959) studied the problem of how to aggregate the items in contingency tables in order to reduce the storage requirements for the tables. Also, Good(1963) introduced a measure of interaction between items by the use of Fourier transform to the probabilities of the items to treate the above problem. These studies were extended by Kullback(1968), Ku & Kullback(1968 & 1969), and Simon(1973). The notion of interaction or correlation between items were, in the most general aspect, treated by Roseboom(1968). Fukao(1971 & 1972) considered the item (state variable) decomposition and aggregation problems for stochastic dynamical

systems by introducing a measure of correlation between items in connection with the stochastic dynamical properties of the systems. These studies , however, are not concerened with the item aggregation problems for contingency tables which represent discrimination information such as in responce matrices.

By trial and error methods, Chow(1966) and Chow & Liu(1968) considered such problems to construct effective Bayesian pattern recognition systems. The measure of correlation offers a theoretical base for the above construction.

Let us compare the case in which two items X_i and X_j are simultaneously observed with the case where they are observed separately. The corresponding response tables are shown in Table 3.4.1(i) & (ii). The quantification ρ in the former case is $\rho((X_i,X_j))$ and in the latter case is $\rho(X_i) + \rho(X_j)$. The difference between them, $\rho((X_i,X_j)) \quad \rho(X_i) \quad \rho(X_j)$, represents the effect of aggregating two items X_i and X_j into one item (X_i,X_j). Therefore, the mean effect can be evaluated by the following quantity.

$$
\begin{aligned}
C_R(X_i,X_j) &= E_{(X_i,X_j)}[\{\rho((X_i,X_j)) \quad \rho(X_i) \quad \rho(X_j)\}\cdot d((X_i,X_j))\ ] \\
&\quad E_{(X_i,X_j)}[\rho((X_i,X_j))\cdot d((X_i,X_j))] \\
&\quad E_{X_i}[\rho(X_i)\cdot d(X_i)] \quad E_{X_j}[\rho(x_j)\cdot d(x_j)] \\
&= S_R((X_i,X_j)) \quad S_R(X_i) - S_R(X_j). \qquad (3.4.1)
\end{aligned}
$$

If the absolute value of the above measure is large, then the items have considerable degree of correlation and the response matrix (table) made by merging the item X_i with the item X_j is supposed to provide a high discrimination rate compared to the original response matrix in which the items X_i and X_j are set separately. From the definition, we have

$$
\begin{aligned}
C_R(X_i,X_j) \triangleq \sum_{X_i,X_j} \{ & P(X_i,X_j,R) \ln \frac{P(X_i,X_j \mid R)}{P(X_i,X_j \mid \bar{R})} \\
& + P(X_i,X_j,\bar{R}) \ln \frac{P(X_i,X_j \mid \bar{R})}{P(X_i,X_j \mid R)} \} \quad S_R(X_i) \quad S_R(X_j). \qquad (3.4.2)
\end{aligned}
$$

From theorem 3.2.2, C_R can be rewritten as follows.

THEOREM 3.4.1 :

$$C_R(X_i,X_j) \quad S_R(X_i \mid X_j) \quad S_R(X_i)$$

$$= S_R(X_j \mid X_i) \quad S_R(X_j), \tag{3.4.3}$$

where $S_R(X_i \mid X_j)$ and $S_R(X_j \mid X_i)$ are the conditional mean information intensities defined by (3.2.22) and (3.2.23). Particularly, when X_i and X_j are stochastically independent in each type R or $\bar{R}$, we have

$$C_R(X_i,X_j) \quad 0 \tag{3.4.4}$$

If X_i and X_j are not stochastically independent, then $C_R(X_i,X_j)$ may be positive or negative; which is the different feature of measure $C_R(X_i,X_j)$ from Shannon's mutual information between items X_i and X_j.

The relation between the measure of correlation and Shannon's mutual information

Let $X_i(R)$ and $X_i(\bar{R})$ represent the random variables for item X_i under the sample type R and $\bar{R}$, respectively, i.e., $X_i(R)$ has the probability law $P(X_i \mid R)$, and $X_i(\bar{R})$ has $P(X_i \mid \bar{R})$. We will show that the measure of correlation $C_R(X_i,X_j)$ has a relationship with Shannon's mutual information measures $M(X_i(R), X_j(R))$ and $M(X_i(\bar{R}), X_j(\bar{R}))$.

From the equalities that

$$P(X_i,X_j,R) + P(X_i,X_j,\bar{R}) = P(X_i,X_j) \tag{3.4.5}$$

$$P(X_i,R) + P(X_i,\bar{R}) = P(X_i), \tag{3.4.6}$$

we have

$$\begin{aligned} C_R(X_i,X_j) = {} & 2P(R)\{\sum_{X_i,X_j} P(X_i,X_j \mid R) \ \ln P(X_i \mid X_j,R) \\ & \sum_{X_i} P(X_i \mid R) \ \ln P(X_i \mid R)\} \\ & + 2P(\bar{R})\{\sum_{X_i,X_j} P(X_i,X_j \mid \bar{R}) \ \ln P(X_i \mid X_j,\bar{R}) \\ & \sum_{X_i} P(X_i \mid \bar{R}) \ \ln P(X_i \mid \bar{R})\} \\ & + \sum_{X_i} P(X_i) \ \ln P(X_i \mid R) + \sum_{X_i} P(X_i) \ \ln P(X_i \mid \bar{R}) \\ & \sum_{X_i,X_j} P(X_i,X_j) \ \ln P(X_i \mid X_j,R) \end{aligned}$$

$$\sum_{X_i,X_j} P(X_i,X_j) \ln P(X_i \mid X_j,\bar{R}). \tag{3.4.7}$$

Also, from the equalities that

$$P(X_i,X_j \mid R) = P(X_i \mid X_j,R)\cdot P(X_j \mid R),$$
$$P(X_i,X_j \mid \bar{R}) = P(X_i \mid X_j,\bar{R})\cdot P(X_j \mid \bar{R}), \tag{3.4.8}$$

we obtain

$$C_R(X_i,X_j) = 2\{P(R)\cdot M(X_i(R),X_j(R)) + P(\bar{R})\cdot M(X_i(\bar{R}),X_j(\bar{R}))\}$$
$$+ h(X_i,X_j ; R,\bar{R}), \tag{3.4.9}$$

where

$$M(X_i(R),X_j(R)) \triangleq H(X_i(R)) \quad H(X_i(R) \mid X_j(R))$$
$$= H(X_i \mid R) \quad H(X_i(R) \mid X_j(R)), \tag{3.4.10}$$

$$M(X_i(\bar{R}),X_j(\bar{R})) \triangleq H(X_i(\bar{R}) \quad H(X_i(\bar{R}) \mid X_j(\bar{R})), \tag{3.4.11}$$

$$h(X_i,X_j ; R,\bar{R}) \triangleq h_0(X_i,X_j ; R) + h_1(X_i,X_j ; \bar{R}), \tag{3.4.12}$$

$$h_0(X_i,X_j ; R) \triangleq \sum_{X_i,X_j} P(X_i,X_j) \ln \frac{P(X_i \mid R)\cdot P(X_j \mid R)}{P(X_i,X_j \mid R)}, \tag{3.4.13}$$

$$h_1(X_i,X_j ; \bar{R}) \triangleq \sum_{X_i,X_j} P(X_i,X_j) \ln \frac{P(X_i \mid \bar{R})\cdot P(X_j \mid \bar{R})}{P(X_i,X_j \mid \bar{R})}. \tag{3.4.14}$$

In order to interpret the meaning of the measures h_0 and h_1, we introduce the hypothetical sample types R' and $\bar{R}'$ as follows:

$$P(X_i,X_j \mid R') = P(X_i \mid R)\cdot P(X_j \mid R), \tag{3.4.15}$$

$$P(X_i,X_j \mid \bar{R}') = P(X_i \mid \bar{R})\cdot P(X_j \mid \bar{R}). \tag{3.4.16}$$

Namely, the types R' and $\bar{R}'$ have the properties that the items X_i and X_j are stochastically independent and the marginal probabilities $P(X_i \mid R')$, $P(X_j \mid R')$, $P(X_i \mid \bar{R}')$, and $P(X_j \mid \bar{R}')$ are the same as $P(X_i \mid R)$, $P(X_j \mid R)$, $P(X_i \mid \bar{R})$, and $P(X_j \mid \bar{R})$, respectively. Then from the definition (3.2.30), h_0 and h_1 are interpreted as the quantified means as follows:

$$h_0(X_i,X_j ; R) = I(R' \mid R ; P(X_i,X_j)) \tag{3.4.17}$$

$$h_1(X_i, X_j ; \bar{R}) \quad I(\bar{R}' \mid \bar{R} \; ; P(X_i, X_j)) \tag{3.4.18}$$

Hence, the quantity h_0 (h_1) represents the degree of belongingness of actual random variable (population) (X_i, X_j) to type $R'(\bar{R}')$ compared to type $R(\bar{R})$.

Equation (3.4.9) says that the measure of correlation $C_R(X_i, X_j)$ between items X_i and X_j is twice the weighted average of Shannon's mutual information measures $M(X_i(R), X_j(R))$ and $M(X_i(\bar{R}), X_j(\bar{R}))$ with the *a priori* probabilities $P(R)$ and $P(\bar{R})$ plus the quantified means $I(R' \mid R \; ; P(X_i, X_j))$ and $I(\bar{R}' \mid \bar{R} \; ; P(X_i, X_j))$. The first term in (3.4.9), Shannon's mutual information, is nonnegative, whereas the second term h may be positive or negative. Thus, the measure $C_R(X_i, X_j)$ may have a negative value.

Qualitative properties of the measure of correlation

We consider the qualitative properties of $C_R(X_i, X_j)$ using a simple example, which is similar to the example in Section 3.2 except for the stochastic dependence between X_i and X_j as follows: The item $X_i(X_j)$ has two categories i_0 $i_1(j_0$ and $j_1)$, and their probabilities are

$$\begin{aligned}
&P(X_i = i_0 \mid R) \quad P(X_j = j_0 \mid R) = 1\text{-}p, \\
&P(X_i = i_1 \mid R) \quad P(X_j = j_1 \mid R) = p, \\
&P(X_i = i_0 \mid \bar{R}) \quad P(X_j = j_0 \mid \bar{R}) = p, \\
&P(X_i = i_1 \mid \bar{R}) = P(X_j \quad j_1 \mid \bar{R}) = 1\text{-}p.
\end{aligned} \tag{3.4.19}$$

The conditional probabilities between item X_i and X_j are

$$\begin{aligned}
&P(X_i = i_0 \mid X_j \quad j_0, R) = P(X_i = i_1 \mid X_j = j_1, \bar{R}) = 1 - p + \alpha p, \\
&P(X_i = i_1 \mid X_j \quad j_1, R) = P(X_i = i_0 \mid X_j \quad j_0, \bar{R}) \quad P + (1 - p)\alpha, \\
&P(X_i = i_0 \mid X_j \quad j_1, R) = P(X_i = i_1 \mid X_j \quad j_0, \bar{R}) \quad p(1 - \alpha), \\
&P(X_i = i_1 \mid X_j = j_0, R) = P(X_i = i_0 \mid X_j = j_1, \bar{R}) \quad (1 - p)(1 - \alpha).
\end{aligned} \tag{3.4.20}$$

Namely, the joint probabilities are

$$\begin{aligned}
&P(X_i = i_0, \; X_j \quad j_0 \mid R) \quad P(X_i = j_1, \; x_j = j_1 \mid \bar{R}) = (1 - p)^2 + p(1 - p)\alpha, \\
&P(X_i = i_1, \; X_j \quad j_1 \mid R) \quad P(X_i = j_0, \; X_j \quad j_0 \mid \bar{R}) = p^2 + p(1 - p)\alpha,
\end{aligned}$$

$$P(X_i = i_1,\ X_j = j_0 \mid R) = P(X_i = i_0,\ X_j = j_1 \mid R)$$

$$P(X_i \quad i_1,\ X_j = j_0 \mid \bar{R})$$

$$= P(X_i = i_0,\ X_j \quad j_1 \mid \bar{R})$$

$$= p(1 - p)(1 - \alpha). \tag{3.4.21}$$

These probability laws are illustrated in Fig. 3.4.1.

The parameter α determines the degree of correlation. The range of α is set as

$$\left(\frac{p}{1 - p}\right) \leq \alpha \leq 1. \tag{3.4.22}$$

The case $\alpha = 0$ corresponds to the case where X_i and X_j are stochastically independent in both types R and $\bar{R}$. The case $\alpha = 1$ or $\alpha = (p / (1 - p))$ corresponds to the case of complete dependence, for we have

$$\text{if } \alpha = 1, \text{ then } P(X_i = i_k \mid X_j = j_k,\ R) = P(X_i = i_k \mid X_j = j_k,\ \bar{R}) = 1,$$

$$\text{for } k = 0,\ 1, \tag{3.4.23}$$

and

$$\text{if } \alpha = \frac{p}{1-p}, \text{ then } P(X_i - i_k \mid X_j = j_{1-k},\ R) = P(X_i \quad i_k \mid X_j = j_{1-k},\ \bar{R})$$

$$= 1, \quad \text{for } k = 0,\ 1. \tag{3.4.24}$$

In this example, the mutual information $M(X_i(R),\ X_j(R))$ and $M(X_i(\bar{R}),\ X_j(\bar{R}))$ coincide with each other, and hence

$$P(R)M(X_i(R),\ X_j(R)) + P(\bar{R})M(X_i(\bar{R}),\ X_j(\bar{R}))$$

$$= M(X_i(R),\ X_j(R))$$

$$M(X_i(\bar{R}),\ X_j(\bar{R})). \tag{3.4.25}$$

We denote the above quantity by $M(p,\alpha)$. Also, in this case, $C_R(X_i,X_j)$ is not dependent on the *a priori* probabilities of R and $\bar{R}$ and is denoted by $C_R(p, \alpha)$. From (3.4.19) and (3.4.20), these measures are calculated as follows:

$$C_R(p,\alpha) = (1 \quad 2p)\left(\ln \frac{1 + \frac{p}{1 - p}\alpha}{1 + \frac{1 - p}{p}\alpha}\right), \tag{3.4.26}$$

$$M(p,\alpha) \quad p(p + \alpha - \alpha p) \ln \left(1 + \frac{1 - p}{p}\alpha\right)$$

$$+ (1-p)(1-p+\alpha p)\ln\left(1+\frac{p}{1-p}\alpha\right)$$

$$+ 2p(1-p)(1-\alpha)\ln(1-\alpha). \tag{3.4.27}$$

If $p \neq \frac{1}{2}$, i.e.,

$$\begin{aligned} P(X_i = i_0 \mid X_j = j_0, R) &= P(X_i = i_0 \mid X_j = j_0, \bar{R}) \\ &= P(X_i = i_1 \mid X_j = j_1, R) \\ &= P(X_i = i_1 \mid X_j = j_1, \bar{R}) \\ &= \frac{1}{2}(1+\alpha), \end{aligned} \tag{3.4.28}$$

$$\begin{aligned} P(X_i = i_0 \mid X_j = j_1, R) &= P(X_i = i_0 \mid X_j = j_1, \bar{R}) \\ &= P(X_i = i_1 \mid X_j = j_0, R) \\ &= P(X_i = i_1 \mid X_j = j_0, \bar{R}) \\ &= \frac{1}{2}(1-\alpha), \end{aligned} \tag{3.4.29}$$

then $C_R = 0$ for arbitrary values of α. On the other hand, Shannon's mutual information M is not equal to 0 provided that $\alpha \neq 0$ ($\alpha = 0$ corresponds to the case where X_i and X_j are stochastically independent). Equations (3.4.28) and (3.4.29) mean that the sample types R and $\bar{R}$, resp., have no difference from R' and $\bar{R}'$ in the following sense:

$$\begin{aligned} \rho((X_i, X_j)) &\triangleq \ln \frac{P(X_i, X_j \mid R)}{P(X_i, X_j \mid \bar{R})} \\ &= \ln \frac{P(X_i \mid R)\cdot P(X_j \mid R)}{P(X_i \mid \bar{R})\cdot P(X_j \mid \bar{R})} \\ &= \rho(X_i) + \rho(X_j) \\ &\triangleq \rho(X_i, X_j). \end{aligned} \tag{3.4.30}$$

This indicates that, in the case $p = \frac{1}{2}$, the quantification $\rho(X_i, X_j)$ is not altered when the items X_i and X_j are aggregated into one item, and hence any discrimination rule based on the quantification is not affected by the aggregation. The property that $C_R = 0$ for arbitrary values of α is in accordance with

the above fact. From (3.4.26), it is easy to see that C_R is a decreasing function of α when $0 \le p \le \frac{1}{2}$ and an increasing function of α under $1 \ge p \ge \frac{1}{2}$. Particularly,

$$\begin{aligned} &\text{if } 0 \le p \le \frac{1}{2}, \quad \text{then } C_R < 0 \quad \text{for } \alpha > 0 \\ &\qquad\qquad\qquad\qquad\quad > 0 \quad \text{for } \alpha < 0, \\ &\text{if } 1 \ge p \ge \frac{1}{2}, \quad \text{then } C_R > 0 \quad \text{for } \alpha > 0 \\ &\qquad\qquad\qquad\qquad\quad < 0 \quad \text{for } \alpha < 0. \end{aligned} \tag{3.4.31}$$

On the other hand,

$$M \ge 0 \quad \text{for any } \alpha. \tag{3.4.32}$$

Fig. 3.4.2 shows the values of C_R and M for various values of α. There is also an essential difference between C_R and M as follows: In the example, it can be readily seen that

$$C_R \to \infty \quad \text{when } \alpha \downarrow -\frac{p}{1-p}, \tag{3.4.33}$$

whereas

$$M \le \ln 2. \tag{3.4.34}$$

Let us consider the next two typical cases as shown in Fig.3.4.3. In the figure, the upper one corresponds to the case $C_R > 0$, and the lower one corresponds to the case $C_R < 0$.

In the figure, ρ means $\rho((X_i,X_j))$ and ρ' means $\rho(X_i) + \rho(X_j)$, i.e., ρ is the quantification for aggregated item (X_i, X_j) and ρ' is the quantification for separate items X_i and X_j.

$$\rho = \ln \frac{P(X_i,X_j \mid R)}{P(X_i,X_j \mid \bar{R})}. \tag{3.4.35}$$

$$\rho' = \ln \frac{P(X_i \mid R)}{P(X_i \mid \bar{R})} + \ln \frac{P(X_j \mid R)}{P(X_j \mid \bar{R})}. \tag{3.4.36}$$

In the upper case ($C_R > 0$), the variation of ρ is dominant over that of ρ'. On the other hand, in the lower case ($C_R < 0$), the variation of ρ' is dominant over that of ρ. Hence the sign of C_R represents the type of correlation between items.

As shown in Fig. 3.4.2, the absolute value of $C_R(X_i, X_j)$ increases along with the departure of α from the origin 0. Thus, if the absolute value is large, then items X_i and X_j have a considerable degree of correlation and the response matrix (table) made by merging item X_i with item X_j should provide a high discrimination rate compared to the original response matrix in which the items X_i and X_j are set separetely.

3.4.2 Optimum Category Aggregation Method based on the Measure of Discriminability

In this section, we consider the optimum method of category aggregation in each item for response matrices. In the sequel, we assume that items X_1, X_2, $\cdots$, and X_n are stochastically independent(dependent cases were discussed in the previous section). As mentioned in Chapter 1, category aggregation means to merge some category types into a new category type.

Each sample $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ is dichotomized into sample type R or $\bar{R}$ according to the quantification $\rho(X_1)$, $\rho(X_2)$, $\cdots$, and $\rho(X_n)$ as discussed in Sections 3.2 and 3.3. When a category aggregation is done for item X_i, the quantification $\rho(X_i)$ is altered, which yields the change of the dichotomization rule. In general, it seems to be obvious that if category aggregation in an item is a refinement of the original aggregation, then the amount of discrimination information of the item increases. If this is the case, there is no problem in studying category aggregation methods. However, if we take note of the fact that, in the usual response matrices, the response probabilities are given or calculated statistically, i.e., based on some data given by *a priori* experiments, then the above statement leaves room for consideration. Namely, if the category aggregation is refined, then the accuracy of each response probability decreases due to the deficiency of the data. Hence, the optimum category aggregation should be determined by taking account of the above accuracy of the response probabilities.

Similar kinds of problems in numerical taxonomy were discussed by Lance & Williams(1966), Wallace & Boulton(1968), and Boulton & Wallace(1969 & 1970).

Based on the above response probabilities, each input(item) X_i is quantified by $\rho(X_i)$ which yields the aforementioned dichotomization rule.

In Section 3.4.2.1, we consider the above problem in a more general framework, where the aggregation of category types is regarded as the compression of discrimination information and the above data is called *a priori information*, and also the item(random variable X_i) is called *input information*(input random

variable)(cf. Katai & Iwai[1974]). In Section 3.4.2.2, the optimum category aggregation problems are discussed based on the method in Section 3.4.2.1(cf. Katai, Imanaga, & Iwai[1972a]). In Section 3.4.2.3, we consider the case where the categories in an item constitute a continuum, and the properties of the optimum category aggregation are examined(cf. Katai, Imanaga, & Iwai[1972b & 1974]).

3.4.2.1 General Theory of Compression of Discrimination Information and the Optimum Method for Discrimination Information Compression

First of all, we take note of the representation of information as *partitions* on probability spaces and the representation of information compression(aggregation) as *subpartitions*. Next, we introduce the notion of *discrimination space* and also that of *discrimination state*. Namely, the discrimination for an input information X is done based on the input information X and *a priori* information Y, hence the triplet consisting of X, Y, and the true sample type plays an essential role in the discrimination for X. We call the triplet the *discrimination state* and also call the probability space to which the discrimination states belong the *discrimination(probability) space*(cf. Katai & Iwai[1974]).

Moreover, we note that the compression of input information X corresponds to a partition on the above discrimination space and the discrimination rate is uniquely determined by the partition. As it is difficult to treat the discrimination rate analytically, we then introduce an information theoretical evaluation method which is clarified as being equal to the measure, the mean information intensity introduced in Section 3.2. By taking note of the *lattice structure* of information and by the use of the above measure, we examine the optimum method of discrimination information compression(aggregation).

Lattice structure of information and the notion of information compression

In the framework of probability theory, a piece of information is represented as a partition on a probability space in the following way: Let Z be a random variable on a probability space (Ω, A, α) and its domain be Z. To observe a sample value of Z brings about the introduction of the next partition $\mathbf{p}$ on Ω (cf. Marchak & Miyasawa[1968]).

$$\omega \equiv \omega'(\mathbf{p}) \overset{\text{def.}}{\Longleftrightarrow} Z(\omega) = Z(\omega'), \text{ for } \omega, \omega' \in \Omega. \tag{3.4.37}$$

The atom containing ω under the partition $\mathbf{p}$ is denoted by $\mathbf{p}(\omega)$, i.e.,

$$\mathbf{p}(\omega) \triangleq \{\omega' \mid \omega' \equiv \omega(\mathbf{p})\}(= Z^{-1}(Z(\omega))). \tag{3.4.38}$$

We denote the set of the atoms of $\mathbf{p}$ by $\phi(\mathbf{p})$, i.e.,

$$\phi(\mathbf{p}) \triangleq \{A \subset \Omega \mid A = \mathbf{p}(\omega),\ \omega \in \Omega\}. \tag{3.4.39}$$

Also, the σ-algebra generated by $\phi(\mathbf{p})$ is denoted by $\sigma(\mathbf{p})$. When a value $z \in Z$ is observed as a sample value of Z, we only know the true value of probablility parameter ω within the coarseness of $\phi(\mathbf{p})$, i.e., we only know that true ω is contained in $Z^{-1}(z)(\in \phi(\mathbf{p}))$. Thus the information contained in Z can be identified with the partition $\mathbf{p}$. When Z, the domain of Z, is a finite set, i.e., when $\#(\phi(\mathbf{p})) < \infty$, then $\mathbf{p}$ is called a *finite* partition. In the sequel, we confine our discussions to the above case to avoid intricate mathematical descriptions.

When the above aspect of information is adopted, then the compression of information is regarded as a subpartition of $\mathbf{p}$ as discussed below. Let $\mathbf{c}$ be a partition on Z. Then the observation of the sample value of Z with the accuracy(the coarseness) of $\mathbf{c}$, i.e., to know the element(atom) of $\mathbf{c}$ which contains the actual sample value z of Z, is represented by the following partition $\mathbf{p}[\mathbf{c}]$:

$$\omega \equiv \omega'(\mathbf{p}[\mathbf{c}]) \overset{\text{def.}}{\Longleftrightarrow} Z(\omega) \equiv Z(\omega')(\mathbf{c}), \tag{3.4.40}$$

where $Z(\omega) \equiv Z(\omega')(\mathbf{c})$ means that both $Z(\omega)$ and $Z(\omega')$ belong to the same element of $\mathbf{c}$. It is easy to see that

$$\mathbf{p}(\omega) \subset \mathbf{p}[\mathbf{c}](\omega) \quad \text{for any } \omega \in \Omega. \tag{3.4.41}$$

Hence, we have

$$\sigma(\mathbf{p}) \supset \sigma(\mathbf{p}[\mathbf{c}]). \tag{3.4.42}$$

Namely, the partition $\mathbf{p}[\mathbf{c}]$ is a subpartition of $\mathbf{p}$.

In general, if we introduce a binary relation "$\geq$" on the set $\Pi(\Omega)$ of the partitions on Ω as follows:

$$\mathbf{p}_1 \geq \mathbf{p}_2 \overset{\text{def.}}{\Longleftrightarrow} \sigma(\mathbf{p}_1) \supset \sigma(\mathbf{p}_2), \tag{3.4.43}$$

then $\geq$ is a (partial) ordering on $\Pi(\Omega)$ and induces a lattice structure into $\Pi(\Omega)$ (for details, refer to Hartmanis[1960], Hartmanis & Sterns[1966], and Iwamura [1966]). Let Z_1 and Z_2 be two random variables on (Ω, A, α) corresponding to partitions $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively. Then the information given by observing the sample values of both Z_1 and Z_2 is represented as the next partition $\mathbf{p}_3$ on Ω.

$$\begin{aligned} \omega \equiv \omega'(\mathbf{p}_3) &\overset{\text{def.}}{\Longleftrightarrow} Z_1(\omega) = Z_1(\omega') \text{ and } Z_2(\omega) \quad Z_2(\omega') \\ &\Longleftrightarrow \omega \equiv \omega'(\mathbf{p}_1) \text{ and } \omega \equiv \omega'(\mathbf{p}_2), \end{aligned} \tag{3.4.44}$$

which is equivalent to

$$\sigma(\mathbb{P}_3) = \{A_1 \cap A_2 \mid A_1 \in \sigma(\mathbb{P}_1),\ A_2 \in \sigma(\mathbb{P}_2)\}$$

; the minimum σ-algebra containing both $\sigma(\mathbb{P}_1)$ and $\sigma(\mathbb{P}_2)$.

(3.4.45)

In contrast with (3.4.45), the information which is common to both Z_1 and Z_2 is represented by the partition $\mathbb{P}_4$ associated with the following σ-algebra.

$$\sigma(\mathbb{P}_4) = \{A \mid A \in \sigma(\mathbb{P}_1) \text{ and } A \in \sigma(\mathbb{P}_2)\}$$

$$\sigma(\mathbb{P}_1) \cap \sigma(\mathbb{P}_2)$$

; the maximum σ-algebra contained in both $\sigma(\mathbb{P}_1)$ and $\sigma(\mathbb{P}_2)$.

(3.4.46)

Let the partitions $\mathbb{P}_3$ and $\mathbb{P}_4$ be denoted as

$$\mathbb{P}_3 = \mathbb{P}_1 \vee \mathbb{P}_2 \quad \text{and} \quad \mathbb{P}_4 \quad \mathbb{P}_1 \wedge \mathbb{P}_2. \tag{3.4.47}$$

Then "$\vee$" and "$\wedge$" correspond to the *join* and the *meet* operations on the above lattice $\Pi(\Omega)$, respectively.

We next introduce the notion of *product partition* as follows: Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be partitions on Ω_1 and Ω_2, respectively. Then the next partition $\mathbb{P}_1 \times \mathbb{P}_2$ on $\Omega_1 \times \Omega_2$ is called the product partition of $\mathbb{P}_1$ and $\mathbb{P}_2$.

$$(\omega_1, \omega_2) \equiv (\omega_1', \omega_2')(\mathbb{P}_1 \times \mathbb{P}_2)$$

$$\overset{\text{def.}}{\Longleftrightarrow} \omega_1 \equiv \omega_1'(\mathbb{P}_1) \text{ and } \omega_2 \equiv \omega_2'(\mathbb{P}_2) \tag{3.4.48}$$

We denote the set of the product partitions on $\Omega_1 \times \Omega_2$ by $\Pi(\Omega_1 \times \Omega_2)$. The lattice operations $\vee$ and $\wedge$ on $\Pi(\Omega_1 \times \Omega_2)$ are given by

$$(\mathbb{P}_1 \times \mathbb{P}_2) \vee (\mathbb{P}_1' \times \mathbb{P}_2') = (\mathbb{P}_1 \vee \mathbb{P}_1') \times (\mathbb{P}_2 \vee \mathbb{P}_2'),$$

$$(\mathbb{P}_1 \times \mathbb{P}_2) \wedge (\mathbb{P}_1' \times \mathbb{P}_2') \quad (\mathbb{P}_1 \wedge \mathbb{P}_2') \times (\mathbb{P}_2 \wedge \mathbb{P}_2'),$$

$$\text{for } \mathbb{P}_1, \mathbb{P}_1' \in \Pi(\Omega_1) \text{ and } \mathbb{P}_2, \mathbb{P}_2' \in \Pi(\Omega_2).$$

(3.4.49)

We denote by $\mathbf{1}_\Omega$ the maximum element of $\Pi(\Omega)$ and by $\mathbb{0}_\Omega$ the minimum element of $\Pi(\Omega)$, i.e.,

$$\mathbf{1}_\Omega(\omega) = \omega, \quad \mathbb{0}_\Omega(\omega) = \Omega, \text{ for any } \omega \in \Omega. \tag{3.4.50}$$

We call $\mathbf{1}_\Omega$ and $\mathbf{0}_\Omega$ *1-partition* and *0-partition* on Ω, respectively.

Introduction of discrimination space and representation of compression of discrimination information as a partition on discrimination space

We consider the following statistical hypothesis testing problem. Suppose that there exists an unknown parameter(random varaible) θ whose domain is $\Theta = \{\theta_0, \theta_1\}$ and we seek the true value of θ. That is to say, we consider the problem of testing which of the hypotheses H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$ is true. In the usual hypothesis testing problems, there exists another random variable X (which we call input random variable) whose probability law depends on the value of θ and the decision is made by knowing the sample value x of X. More precisely, the decision rule is made based on the probability measure $\nu_X(\cdot \mid \theta)$ on sample (probability) space $(X, \mathcal{C}_X)$ of X as follows:

$$\nu_X(x \mid \theta) = \text{Prob.}(X = x \mid \theta) \quad \text{for } x \in X, \ \theta = \theta_0, \theta_1. \tag{3.4.51}$$

In Bayesian hypothesis testing, *a priori* probability measure λ is introduced on the parameter space $(\Theta, \mathcal{B}_\Theta)$, i.e., the parameter space is considered as the parameter probability space $(\Theta, \mathcal{B}_\Theta, \lambda)$, where $\mathcal{B}_\Theta$ consists of ϕ, $\{\theta_0\}$, $\{\theta_1\}$, and Θ. The decision of the true value of θ is made based on the following probability measure κ on the product space $(\Sigma_{(\Theta, X)}, \mathcal{F}_{(\Theta, X)}) \triangleq (\Theta, \mathcal{B}_\Theta) \times (X, \mathcal{C}_X)$.

$$\kappa(\sigma) = \nu_X(x \mid \theta) \cdot \lambda(\theta), \quad \sigma \quad (\theta, x) \in \Sigma_{(\Theta, X)}. \tag{3.4.52}$$

The Bayesian decision rule in this case is given as

$$\left.\begin{array}{l} \text{if } L(x) > 1, \text{ then accept } H_0\text{: } \theta \quad \theta_0 \\ \text{if } L(x) < 1, \text{ then accept } H_1\text{: } \theta \quad \theta_1 \end{array}\right\}, \tag{3.4.53}$$

where

$$L(x) \triangleq \frac{\kappa(\theta_0, x)}{\kappa(\theta_1, x)} . \tag{3.4.54}$$

As discussed previously, the compression of information given by X is represented by a partition $\mathbf{c}_X$ on X. When the sample value x of X is observed in the accuracy of $\mathbf{c}_X$, then the decision rule becomes

$$\left.\begin{array}{l} \text{if } L(x : \mathbf{c}_X) > 1, \text{ then } \theta = \theta_0 \\ \text{if } L(x : \mathbf{c}_X) < 1, \text{ then } \theta = \theta_1 \end{array}\right\}, \tag{3.4.55}$$

where

$$L(x : \mathbf{c}_X) \triangleq \frac{\kappa(\{\theta_0\} \times \mathbf{c}_X(x))}{\kappa(\{\theta_1\} \times \mathbf{c}_X(x))} \quad \frac{\sum_{x' \in \mathbf{c}_X(x)} \kappa(\theta_0, x')}{\sum_{x' \in \mathbf{c}_X(x)} \kappa(\theta_1, x')} \tag{3.4.56}$$

In this section, as aforementioned, we consider the case where ν_X is unknown and is estimated based on the sample value y of another random variable Y (which is called *a priori* information) whose probability structure depends on the true value of θ. That is, in (3.4.52), the conditional probabilities $\{\nu_X(\cdot \mid \theta)\}_{\theta \in \Theta}$ are replaced by the estimated values $\{\hat{\nu}_X(\cdot \mid \theta, y)\}_{\theta \in \Theta}$ given by the sample value y of Y.

When the sample value x of X is observed in the accuracy of partition $\mathbf{c}_X$(on X), then the corresponding decision rule is given as

$$\hat{L}(x : \mathbf{c}_X, y) \triangleq \frac{\hat{\nu}_X(\mathbf{c}_X(x) \mid \theta_0, y) \cdot \lambda(\theta_0)}{\hat{\nu}_X(\mathbf{c}_X(x) \mid \theta_1, y) \cdot \lambda(\theta_1)}, \tag{3.4.57}$$

if $\hat{L}(x : \mathbf{c}_X, y) > 1$, then $\theta = \theta_0$, and

if $\hat{L}(x : \mathbf{c}_X, y) < 1$, then $\theta \quad \theta_1$. (3.4.58)

There are many methods to obtain $\hat{\nu}_X$(the estimation of ν_X), and they are discussed in the general framework of probability estimation methods, hence we do not refer to them in detail and presume that a method is already given beforehand(cf. Cover[1972] and Noguchi, Tomita, & Watanabe[1972]; in Section 3.4.2.2, we discuss the case of estimation based on the relative frequency of occurrence of each category type in the *a priori* information(data)).

As discussed above, the decision rule depends on the sample value y of the *a priori* information Y. The probability structure of Y is represented by the (sample) probability measure μ_Y on the sample space $(\mathcal{Y}, \mathcal{D}_Y)$ of Y as follows:

$$\mu_Y(y) \quad \text{Prob.}(Y = y) \quad \text{for } y \in \mathcal{Y}. \tag{3.4.59}$$

Accordingly, the joint probability structure of the parameter θ, the input random variable X and the *a priori* information Y is given by the (joint) probability measure τ on the product space:

$$(\Xi_{(\Theta,\ X,\ Y)},\ \mathcal{E}_{(\Theta,\ X,\ Y)}) \triangleq (\Theta,\ \mathcal{B}_{\Theta}) \times (X,\ \mathcal{C}_X) \times (Y,\ \mathcal{D}_Y). \tag{3.4.60}$$

In the usual cases, the *a priori* information Y and the input information X are stochastically independent. Hence the joint probability measure τ is given as

$$\tau(\xi) \quad \kappa(\theta,\ x)\cdot\mu_Y(y) = \nu_X(x \mid \theta)\cdot\lambda(\theta)\cdot\mu_Y(y),$$

$$\text{for } \xi - (\theta,\ x,\ y) \in \Xi_{(\Theta,\ X,\ Y)}. \tag{3.4.61}$$

The space $(\Xi_{(\Theta,\ X,\ Y)},\ \mathcal{E}_{(\Theta,\ X,\ Y)})$ contains all the necessary entities of this discrimination problem, i.e., it represents the so-called "universe of discourse", and we call it the *discrimination space* and its element $\xi(\in \Xi_{(\Theta,\ X,\ Y)})$ a *discrimination state* (cf. Katai & Iwai[1974]).

To represent the decision rule given by (3.4.57) and (3.4.58) on the above discrimination space, we introduce another probability measure $\hat{\tau}$ (on the discrimination space) corresponding to $\hat{\nu}_X$ as follows:

$$\hat{\tau}(\xi) \triangleq \hat{\nu}_X(x \mid \theta,\ y)\cdot\lambda(\theta)\cdot\mu_Y(y) \quad \text{for } \xi \in \Xi_{(\Theta,\ X,\ Y)}. \tag{3.4.62}$$

Then (3.4.57) is rewritten as

$$\hat{L}(x : \mathbf{c}_X,\ y) \quad \frac{\hat{\tau}(\{\theta_0\} \times \mathbf{c}_X(x) \times \{y\})}{\hat{\tau}(\{\theta_0\} \times \mathbf{c}_X(x) \times \{y\})}$$

$$\left(= \frac{\sum_{x' \in \mathbf{c}_X(x)} \hat{\tau}(\theta_0,\ x',\ y)}{\sum_{x' \in \mathbf{c}_X(x)} \hat{\tau}(\theta_1,\ x',\ y)} \right). \tag{3.4.63}$$

The decision rule (3.4.58) says that if the input information X is compressed by the partition $\mathbf{c}_X$, then the *a priori* information Y is also compressed by the following partition $\mathbf{d}_Y[\mathbf{c}_X]$ in accordance with the partition $\mathbf{c}_X$. This is because the sample values y's of Y which yield the same value of $\hat{L}(x : \mathbf{c}_X,\ y)$ can be identified with each other, as far as the decision rule (3.4.58) is concerned

$$y \equiv y'(\mathbf{d}_Y[\mathbf{c}_X]) \overset{\text{def.}}{\Longleftrightarrow} \hat{L}(x : \mathbf{c}_X,\ y) = \hat{L}(x : \mathbf{c}_X,\ y') \quad \text{for any } x \in X. \tag{3.4.64}$$

Obviously, there exists a monotonous relation between the partitions $\mathbf{c}_X$ and $\mathbf{d}_Y[\mathbf{c}_X]$, i.e.,

$$\text{if } \mathbf{c}_X \geq \mathbf{c}_X', \text{ then } \mathbf{d}_Y[\mathbf{c}_X] \geq \mathbf{d}_Y[\mathbf{c}_X']. \tag{3.4.65}$$

From the definition (3.4.64), (3.4.63) can be modified as

$$\hat{L}(x : \mathbf{c}_X, y) \quad \frac{\hat{\tau}(\{\theta_0\} \times \mathbf{c}_X(x) \times \mathbf{d}_Y[\mathbf{c}_X](y))}{\hat{\tau}(\{\theta_1\} \times \mathbf{c}_X(x) \times \mathbf{d}_Y[\mathbf{c}_X](y))}$$

$$\left(\frac{\sum_{x' \in \mathbf{c}_X(x),\ y' \in \mathbf{d}_Y[\mathbf{c}_X](y)} \hat{\tau}(\theta_0, x', y')}{\sum_{x' \in \mathbf{c}_X(x),\ y' \in \mathbf{d}_Y[\mathbf{c}_X](y)} \hat{\tau}(\theta_1, x', y')} \right). \tag{3.4.66}$$

The above equation means that when input information X is compressed by the partition $\mathbf{c}_X$, then the information contained in X and Y is represented by the partition $\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]$ on $X \times Y$, where $\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]$ is the product partition as follows(cf. (3.4.48)):

$$(x, y) \equiv (x', y')(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X])$$

$$\Longleftrightarrow x \equiv x'(\mathbf{c}_X) \text{ and } y \equiv y'(\mathbf{d}_Y[\mathbf{c}_X]), \text{ i.e.,}$$

$$\mathbf{c}_X(x) \times \mathbf{d}_Y[\mathbf{c}_X](x, y) = \mathbf{c}_X(x) \times \mathbf{d}_Y[\mathbf{c}_X](y)$$

$$\text{for any } x \in X \text{ and any } y \in Y. \tag{3.4.67}$$

In order to represent the above information(partition) on the discrimination space, we consider the product partition of $\mathbf{1}_\Theta$(1-partition on Θ) and the above partition and denote it by $\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]]$, i.e.,

$$\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]] \triangleq \mathbf{1}_\Theta \times \mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]. \tag{3.4.68}$$

This is a partition on $\Xi_{(\Theta, X, Y)}$, and (3.4.66) can be rewritten as

$$\hat{L}(x : \mathbf{c}_X, y) = \begin{cases} \hat{L}_\Xi(\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]](\xi)) & \text{if } \theta = \theta_0 \\ \hat{L}_\Xi(\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]](\xi))^{-1} & \text{if } \theta = \theta_1, \end{cases} \tag{3.4.69}$$

where

$$\hat{L}_\Xi(E) \triangleq \frac{\hat{\tau}(E)}{\hat{\tau}_c(E)} \quad \text{for any } E \in \mathcal{E}_{(\Theta, X, Y)} \tag{3.4.70}$$

and $\hat{\tau}_c$ is a probability measure on $(\Xi_{(\Theta, X, Y)}, \mathcal{E}_{(\Theta, X, Y)})$ given as follows:

$$\hat{\tau}_c(\theta_i, x, y) \triangleq \tau(\theta_{1-i}, x, y) \quad \text{for } i = 0, 1, x \in X, \text{ and } y \in Y. \tag{3.4.71}$$

By the use of (3.4.69), it follws that when the discrimination state is at ξ (θ_i, x, y), then the Bayesian decision rule (3.4.58) is restated as

$$\hat{L}_{\Xi}(\mathbf{e}_{\Xi}[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]](\xi)) \quad \begin{cases} > 1 & \Rightarrow \text{accept } H_i\colon \theta \quad \theta_i \\ < 1 & \Rightarrow \text{accept } H_{1-i}\colon \theta = \theta_{1-i}. \end{cases} \tag{3.4.72}$$

That is to say, the partition $\mathbf{e}_{\Xi}[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]]$ on the discrimination space plays an essential role in the Bayesian discrimination. From (3.4.72), the region Ξ_c(in $\Xi_{(\Theta, X, Y)}$) in which the discrimination is done correctly is given as

$$\Xi_c = \{\xi \mid \hat{L}_{\Xi}(\mathbf{e}_{\Xi}[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]](\xi)) > 1\}. \tag{3.4.73}$$

There are two methods to measure the size of the above region. One way is to measure it by the (objective) probability measure τ and the resultant size is called the objective discrimination rate and is denoted by $P_c(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X])$. The other is to measure it by the (subjective) probability measure $\hat{\tau}$, and we call the resultant size the subjective discrimination rate and denote it by $\hat{P}_c(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X])$. Namely,

$$\begin{aligned} P_c(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]) &\triangleq \tau(\Xi_c) \\ &= E_{\tau}[s(\hat{L}_{\Xi}(\mathbf{e}_{\Xi}[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]](\xi))] \\ &= \sum_{E \in \phi(\mathbf{e}_{\Xi}[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]])} \tau(E) \cdot s\left(\frac{\hat{\tau}(E)}{\hat{\tau}_c(E)}\right), \end{aligned} \tag{3.4.74}$$

$$\begin{aligned} \hat{P}_c(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]) &\triangleq \hat{\tau}(\Xi_c) \\ &= \sum_{E \in \phi(\mathbf{e}_{\Xi}[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]])} \hat{\tau}(E) \cdot s\left(\frac{\hat{\tau}(E)}{\hat{\tau}_c(E)}\right), \end{aligned} \tag{3.4.75}$$

where s is a discontinuous function on the real axis as follows:

$$s(r) \quad \begin{cases} 1, \text{if } r > 1 \\ 0, \text{if } r < 1. \end{cases} \tag{3.4.76}$$

It should be noted that Bayesian discrimination cannot be done for all the possible values of the pair (x, y). For instance, if the *a priori* information y gives that $\hat{\nu}_X(x \mid \theta_0, y) = \hat{\nu}_X(x \mid \theta_1, y) = 0$, then $\hat{L}$ cannot be calculated and the Bayesian discrimination cannot be done. In other words, if the *a priori* information y gives only insufficient information for the discrimination based on x, then the discrimination by x cannot be carried out. Therefore, the discrimination is done only on the subregion Ξ_a(which we call *admissible region*) of the discrimination space $\Xi_{(\Theta, X, Y)}$. Hence the optimum compression $\mathbf{c}_X^*$ of input information X in the sense of Bayesian discrimination is given by

the following criterion.

$$\tau(\Xi_a \cap \Xi_c) \rightarrow \max \quad \text{w.r.t. } \mathbf{c}_X, \tag{3.4.77}$$

where

$$\tau(\Xi_a \cap \Xi_c) = \sum_{E_a \in \phi_a(\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]])} \tau(E_a) \cdot s\left(\frac{\hat{\tau}(E_a)}{\hat{\tau}_c(E_a)}\right), \tag{3.4.78}$$

$$\phi_a(\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]]) \triangleq \{E_a \mid E_a = E \cap \Xi_a,\ E \in \phi(\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]])\}. \tag{3.4.79}$$

In the above criterion, it should be noted that the partition $\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]]$ is uniquely determined by the partition $\mathbf{c}_X$.

Introduction of a topology into the lattice structure of information and quantitative analysis of the optimum compression of input information

It is difficult, in general, to obtain the optimum compression $\mathbf{c}_X^*$ based on the criterion (3.4.77), because of the complexity of the dependence of the partition $\mathbf{d}_Y[\mathbf{c}_X]$ on $\mathbf{c}_X$ and also of the nonlinearity of the function s defined by (3.4.76). So, we seek the suboptimum compression (partition) $\mathbf{c}_X^{**}$ according to the following considerations. First, we consider the next criterion.

$$\sum_{E_a \in \phi_a[\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y]]} \tau(E_a) \cdot s\left(\frac{\hat{\tau}(E_a)}{\hat{\tau}_c(E_a)}\right) \rightarrow \max \text{ w.r.t. } \mathbf{c}_X \text{ and } \mathbf{d}_Y, \tag{3.4.80}$$

where

$$\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y] = \mathbf{1}_\Theta \times \mathbf{c}_X \times \mathbf{d}_Y, \tag{3.4.81}$$

$$\phi_a(\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y]) = \{E_a \mid E_a = E \cap \Xi_a,\ E \in \phi(\mathbf{e}_\Xi[\mathbf{c}_X \times \mathbf{d}_Y])\}. \tag{3.4.82}$$

In the above criterion, the partition $\mathbf{d}_Y$ on Y is regarded as independent of $\mathbf{c}_X$, and the optimum partition pair $(\mathbf{c}_X, \mathbf{d}_Y)$ is searched for. In the criterion (3.4.77), $\mathbf{d}_Y$ is determined by $\mathbf{c}_X$ and only the optimum partition $\mathbf{c}_X$ is searched for.

Let $(\mathbf{c}_X^o, \mathbf{d}_Y^o)$ be the optimum pair of partitions for the *a priori* information X and the input information Y. The optimum pair $(\mathbf{c}_X^o, \mathbf{d}_Y^o)$ yields the optimum (product) partition $\mathbf{c}_X^o \times \mathbf{d}_X^o$ on $X \times Y$. In general cases, there may exist more than one optimum pairs, hence the set of the optimum partitions $\{\mathbf{c}_X^o \times \mathbf{d}_Y^o\}$ constitute a subset Π_o of the set $\Pi(X \times Y)$ of the product partitions on X and Y.

On the other hand, the optimum partition $\mathbf{c}_X^*$ of the input information and

the corresponding partition $\mathbf{d}_Y[\mathbf{c}_X^*]$ belong to the subset Π_r of $\Pi(X \times Y)$ as follows:

$$\Pi_r \triangleq \{\mathbf{c}_X \times \mathbf{d}_Y \mid \mathbf{d}_Y = \mathbf{d}_Y[\mathbf{c}_X],\ \mathbf{c}_X \in \Pi(X)\}. \tag{3.4.83}$$

The above region Π_r represents the actually available information given by the *a priori* information Y and the input information X. Therefore, we search for the partition $\mathbf{c}_X^{**} \times \mathbf{d}_Y^{**}$ in Π_r which is the most proximate to the optimum region Π_o. The first component $\mathbf{c}_X^{**}$ of $\mathbf{c}_X^{**} \times \mathbf{d}_Y^{**}$ $\mathbf{c}_X^{**} \times \mathbf{d}_Y[\mathbf{c}_X^{**}]$ gives the suboptimum partition.

In the above approach, we introduce a topology into the set $\Pi(X \times Y)$ to measure the proximity. First, we take note of the lattice structure of $\Pi(X \times Y)$. In general, the introduction of a topology into a lattice Π $\{\mathbf{h}\}$ with the property that

$$(\mathbf{h}_1 \wedge \mathbf{h}_2) \vee (\mathbf{h}_1 \wedge \mathbf{h}_3) = \mathbf{h}_1 \wedge (\mathbf{h}_2 \vee (\mathbf{h}_1 \wedge \mathbf{h}_3)) \quad \text{for any } \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3 \in \Pi \tag{3.4.84}$$

is done by the next method.

$$\bar{\Delta}(\mathbf{h}_1, \mathbf{h}_2) \triangleq m(\mathbf{h}_1 \vee \mathbf{h}_2) \quad m(\mathbf{h}_1 \wedge \mathbf{h}_2), \tag{3.4.85}$$

where m is a real valued function satisfying

$$m(\mathbf{h}_1 \vee \mathbf{h}_2) + m(\mathbf{h}_1 \wedge \mathbf{h}_2) \quad m(\mathbf{h}_1) + m(\mathbf{h}_2). \tag{3.4.86}$$

In the above, a lattice satisfying (3.4.84) is called a *modular lattice* and a function m satisying (3.4.86) is called a *modular function*. A typical example of a modular lattice is the Boolear algebra, and probability measures are typical examples of modular functions. When m is a modular function, $\bar{\Delta}$ defined by (3.4.85) is a metric on Π, i.e., it satisfies the axiom of metric(norm), and can be rewritten as

$$\bar{\Delta}(\mathbf{h}_1, \mathbf{h}_2) = (m(\mathbf{h}_1) \quad m(\mathbf{h}_1 \wedge \mathbf{h}_2)) + (m(\mathbf{h}_2) \quad m(\mathbf{h}_1 \wedge \mathbf{h}_2)). \tag{3.4.87}$$

The lattice $\Pi(X \times Y)$ is, in general, not a modular lattice. However, we adopt the above method (3.4.87) to introduce a topology into $\Pi(X \times Y)$, because there is no other appropriate way to introduce a topology into $\Pi(X \times Y)$ and it is possible to give an information theoretical interpretation of the metric $\bar{\Delta}$ by adopting an appropriate function m as shown in the sequel.

It is desirable to adopt a function which is closely related to the criterion (3.4.80) as the function m in (3.4.87). Because the objective probability τ is unknown to us and the discontinuous function s is difficult to

treat, we approximate the criterion (3.4.80) by the following information theoretic measure which is given by substituting the subjective probability $\hat{\tau}$ for τ and the natural logarithmic function for the function s, i.e., we adopt the following measure S as the function m.

$$S(\mathbf{h}) \overset{\Delta}{=} \sum_{E_a \in \phi_a(\mathbf{e}_\Xi[\mathbf{h}])} \hat{\tau}(E_a) \cdot \ln \frac{\hat{\tau}(E_a)}{\hat{\tau}_c(E_a)} \quad \text{for } \mathbf{h} \in \Pi(X \times Y), \tag{3.4.88}$$

where

$$\mathbf{e}_\Xi[\mathbf{h}] \overset{\Delta}{=} \mathbf{1}_\Theta \times \mathbf{h} \quad \text{for } \mathbf{h} \in \Pi(X \times Y). \tag{3.4.89}$$

Instead of the above criterion, Barndorff-Nielsen(1964) introduced a minimax type criterion for similar problems.

The above measure $S(\mathbf{h})$ is the Kullback-Leibler information number evaluating the difference of the probability measure $\hat{\tau}$ from $\hat{\tau}_c$ on the discrimination space $(\Xi_{(\Theta, X, Y)}, \mathcal{E}_{(\Xi, X, Y)})$(more precisely, under the partition $\mathbf{e}_\Xi[\mathbf{h}]$ and the restriction to Ξ_a), and can be interpreted as the amount of discrimination information contained in the partition $\mathbf{h} \in \Pi(X \times Y)$.

Accordingly, the metric $\bar{\Delta}$ on $\Pi(X \times Y)$ is given by

$$\bar{\Delta}(\mathbb{h}_1, \mathbb{h}_2) \overset{\Delta}{=} \Delta(\mathbb{h}_1, \mathbb{h}_1 \wedge \mathbf{h}_2) + \Delta(\mathbb{h}_2, \mathbb{h}_1 \wedge \mathbf{h}_2) \quad \text{for } \mathbb{h}_1, \mathbb{h}_2 \in \Pi(X \times Y), \tag{3.4.90}$$

where

$$\Delta(\mathbf{h}, \mathbf{h}') \overset{\Delta}{=} S(\mathbf{h}) \quad S(\mathbf{h}'). \tag{3.4.91}$$

The difference Δ represents the amount of information loss due to the compression of information(partition) $\mathbb{h}$ into subpartition $\mathbb{h}'$. Also, as clarified later, Δ is represented as the decrease of the mean information intensity introduced in Section 3.2 when applied to the category aggregation problems for response matrices.

From Jensen inequality on convex functions, we have

$$\Delta(\mathbf{h}, \mathbb{h}') \geq 0 \quad \text{for any } \mathbf{h} \text{ and } \mathbf{h}' \text{ such that } \mathbf{h} \geq \mathbf{h}', \tag{3.4.92}$$

$$\Delta(\mathbf{h}, \mathbf{h}') = 0 \quad \text{iff} \quad \frac{\hat{\tau}(E_a)}{\hat{\tau}_c(E_a)} = \frac{\hat{\tau}(E_a')}{\hat{\tau}_c(E_a')} \quad \text{for any } E_a \text{ and } E_a' \text{ such that } E_a \subset E_a',\ E_a \in \phi_a(\mathbf{e}_\Xi[\mathbf{h}]), \text{ and } E_a' \in \phi_a(\mathbf{e}_\Xi[\mathbb{h}']). \tag{3.4.93}$$

The condition part of (3.4.93) is the so-called "sufficiency" condition of information $\mathbf{h}'$ with respect to $\mathbb{h}$(cf. Kullback[1951] and Ghurye[1968]). It

means that the compression from $\mathbb{h}$ into $\mathbb{h}'$ has no effect on the discrimination between θ_0 and θ_1, and hence the above measure Δ(and S) is in pursuance of our intention.

The measure $\bar{\Delta}(\mathbb{h}_1, \mathbb{h}_2)$ of the difference between the partitions(information) $\mathbb{h}_1$ and $\mathbb{h}_2$ is the sum of the amounts of information $\Delta(\mathbb{h}_1, \mathbb{h}_1 \wedge \mathbb{h}_2)$ and $\Delta(\mathbb{h}_2, \mathbb{h}_1 \wedge \mathbb{h}_2)$ needed to obtain information $\mathbb{h}_1$ and $\mathbb{h}_2$ from the information $\mathbb{h}_1 \wedge \mathbb{h}_2$, i.e. the common information in $\mathbb{h}_1$ and $\mathbf{h}_2$.

By the use of the above measure $\bar{\Delta}$, the suboptimum partition $\mathbb{c}_X^{**}$ of input information X is given by the next criterion.

$$d(\mathbf{c}_X^{**} \times \mathbf{d}_Y[\mathbf{c}_X^{**}], \Pi_o) \quad \min_{\mathbf{c}_X \in \Pi(X)} d(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X], \Pi_o), \tag{3.4.94}$$

where

$$d(\mathbf{c}_X \times \mathbf{d}_Y, \Pi_o) \triangleq \min_{\mathbf{c}_X^o \times \mathbf{d}_Y^o \in \Pi_o} \bar{\Delta}(\mathbb{c}_X \times \mathbb{d}_Y, \mathbb{c}_X^o \times \mathbb{d}_Y^o). \tag{3.4.95}$$

Fig. 3.4.4 depicts the above relationships(in the strict sense of the term, $\Pi(X)$ and $\Pi(Y)$ are not totally ordered set, hence they cannot be represented on the lines).

3.4.2.2 Optimum Category Aggregation Method based on the Measure

In this section, we consider the problem of the optimum category aggregation in response matrices(tables) based on the general method introduced in Section 3.4.2.1. We consider the above problem under the supposition that the items in a response table are stochastically independent.

In the general discussions in Section 3.4.2.1, the input information(random variable) X corresponds to the collection of items $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ and, the partition $\mathbf{c}_X$ corresponds to the category aggregation of the items, the values of the parameters θ_0 and θ_1 correspond to the sample types R and $\bar{R}$ respectively, λ corresponds to the *a priori* probabilities of the sample types, ν_X corresponds to the response probabilities, and the *a priori* information (random variable) Y corresponds to the information which is used to construct the response matrix. Hence, it seems to be natural that the optimum category aggregation problem should be discussed for all the items X_1, X_2, $\cdots$, and X_n, simultaneously. However, from the supposition of stochastic independence of the items and also from the additivity of Kullback-Leibler information number for stochastically independent random variables, the measure $S(\mathbf{h})$ defined by (3.4.88) and hence the measures Δ, $\bar{\Delta}$, and d can be calculated itemwise.

Therefore, we can treat the optimum category aggregation problem for each item based on the criterion (3.4.94), separetely, where $\mathfrak{C}_X$ corresponds to the category aggregation of a single item X.

In the usual response table, the estimated response probability $\hat{\nu}_X$ of ν_X is given as the relative frequencies in the finite sample values(in an *a priori* experiment) of the item X, i.e., the *a priori* information Y corresponding to the item X is given as

$$Y \quad (Y_0, Y_1),$$

$$Y_i \quad (X_{i1}, X_{i2}, \cdots, X_{in_i}), \ i = 0, 1, \tag{3.4.96}$$

where X_{ij} is the j^{th} sample of item X under $\theta = \theta_i (i = 0, 1)$ and n_i is the number of the samples in the *a priori* experiment, i.e.,

$$\mu_Y(X_{ij} = x_{ij}, \ j \quad 1, 2, \cdots, n_i, \ i = 0, 1)$$

$$= \prod_{j=1}^{n_0} \nu_X(x_{0j} \mid \theta_0) \cdot \prod_{j=1}^{n_1} \nu_X(x_{1j} \mid \theta_1). \tag{3.4 97}$$

In the above, x_{ij} represents the sample value of X_{ij} for i = 0, 1, and j 1, 2, $\cdots$, n_i. The estimation $\hat{\nu}_X$ of ν_X is given as the relative frequencies in the sample y of Y as follows:

$$\hat{\nu}_X(x \mid \theta_i, y) \triangleq n_i^{-1} \cdot \sum_{j=1}^{n_i} \chi_x(x_{ij}), \quad i \quad 0, 1, \tag{3.4.98}$$

where

$$y = (y_0, y_1),$$

$$y_i \quad (x_{i1}, x_{i2}, \cdots, x_{in_i}) \quad \text{for } i = 0, 1, \tag{3.4.99}$$

$$\chi_x(x') \triangleq \begin{cases} 1 & \text{if } x' \neq x, \\ 0 & \text{if } x' \quad x, \end{cases} \tag{3.4.100}$$

for any x and $x' \in X$.

From (3.4.97), the next relation holds.

$$E_{\mu_Y}[\chi_x(x_{ij})] = \text{Prob.}(X_{ij} \quad x_{ij}) = \nu_X(x \mid \theta_i). \tag{3.4.101}$$

It follows that $\hat{\nu}_X$ is an unbiased estimator of ν_X(cf. Katai, Imanaga, & Iwai [1972a]), i.e.,

$$E_{\mu_Y}[\hat{\nu}_X(x \mid \theta_i, y)] = \sum_{y \in y} \hat{\nu}_X(x \mid \theta_i, y) \cdot \mu_Y(y)$$

$$= \nu_X(x \mid \theta_i). \tag{3.4.102}$$

Also, the above relation is rewritten as

$$\hat{\tau}(\{\theta_i\} \times \{x\} \times Y)\left(= \sum_{y \in Y} \hat{\tau}(\theta_i, x, y)\right)$$

$$= \tau(\theta_i, x) \quad \text{for } i \quad 0, 1. \tag{3.4.103}$$

In the following, we confine ourselves to the case where the response probabilities $\nu_X(\cdot \mid \theta_0)$ and $\nu_X(\cdot \mid \theta_1)$ are *positive* measures on X, i.e.,

$$\nu_X(x \mid \theta_i) > 0 \quad \text{for any } x \in X \text{ and } i = 0, 1. \tag{3.4.104}$$

In this case, $\hat{\nu}_X(x \mid \theta_i, y) \quad 0$ does not mean that $\nu_X(x \mid \theta_i) \quad 0$ but means that that sample y does not offer any information about the value x of X under $\theta \quad \theta_i$. Therefore, we set the admissible region Ξ_a as follows:

$$\Xi_a \triangleq \{\xi = (\theta, x, y) \mid \hat{\nu}_X(x \mid \theta_0, y) > 0 \text{ and } \hat{\nu}_X(x \mid \theta_1, y) > 0\} \tag{3.4.105}$$

That is, $\hat{\tau}$ and $\hat{\tau}_c$ are positive measures on Ξ_a, and $S(\mathbf{h})$ is finite for any partition $\mathbf{h}$ on $\Xi_{(\Theta, X, Y)}$.

Let us search for the optimum region Π_o by noting the unbiasedness of $\hat{\nu}_X$ to ν_X. From (3.4.103) and (3.4.105), we have

$$\hat{\tau}(\mathbf{e}_\Xi[\mathbf{1}_X \times \mathbf{0}_Y](\xi) \cap \Xi_a) = \hat{\tau}(\{\theta_i\} \times \{x\} \times Y \cap \Xi_a)$$

$$\tau(\{\theta_i\} \times \{x\} \times Y)$$

$$= \kappa(\theta_i, x), \quad \text{when } \xi = (\theta_i, x, y) \tag{3.4.106}$$

The above means that the ideal discrimination rule (3.4.53)(which yields the maximum discrimination rate) can be restated in terms of the probability $\hat{\tau}$ as follows:

$$\frac{\hat{\tau}(\mathbf{e}_\Xi[\mathbf{1}_X \times \mathbf{0}_Y](\xi) \cap \Xi_a)}{\hat{\tau}_c(\mathbf{e}_\Xi[\mathbf{1}_X \times \mathbf{0}_Y](\xi) \cap \Xi_a)} \begin{cases} > 1 \implies \text{accept } H_i \\ < 1 \implies \text{accept } H_{1-i}, \end{cases} \quad \text{when } \xi = (\theta_i, x, y) \tag{3.4.107}$$

That is to say, $\mathbf{1}_X \times \mathbf{0}_Y$, the product partition of 1-partition on X and 0-partition on Y, yields the optimum decision rule, hence it is contained in the optimum region Π_o. There may be other members in Π_o, however, those cannot be specified unless the concrete values of ν_X are given to us. Hence, we assume

that

$$\Pi_o = \{\mathbf{1}_X \times \mathbf{0}_Y\} \tag{3.4.108}$$

From (3.4.49) and (3.4.87), (3.4.95) is rewritten as

$$\begin{aligned} d(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X],\ \Pi_o) &= \bar{\Delta}(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X],\ \mathbf{1}_X \times \mathbf{0}_Y) \\ &= \Delta_1(\mathbf{c}_X) + \Delta_2(\mathbf{c}_X), \end{aligned} \tag{3.4.109}$$

where

$$\begin{aligned} \Delta_1(\mathbf{c}_X) &\triangleq \Delta(\mathbf{1}_X \times \mathbf{0}_Y,\ \mathbf{c}_X \times \mathbf{0}_Y) \\ &\quad S(\mathbf{1}_X \times \mathbf{0}_Y) \quad S(\mathbf{c}_X \times \mathbf{0}_Y), \end{aligned} \tag{3.4.110}$$

$$\begin{aligned} \Delta_2(\mathbf{c}_X) &\triangleq \Delta(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X],\ \mathbf{c}_X \times \mathbf{0}_Y) \\ &\quad S(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]) \quad S(\mathbf{c}_X \times \mathbf{0}_Y). \end{aligned} \tag{3.4.111}$$

The suboptimum compression(aggregation) $\mathbf{c}_X^{**}$ indicated by (3.4.94) is given as the compression $\mathbf{c}_X$ which yields the minimum of d defined by (3.4.109).

From (3.4.88) and (3.4.106), the measure Δ_1 is rewritten as

$$\begin{aligned} \Delta_1(\mathbf{c}_X) &\quad \sum_{l=0,1} \sum_{x \in X} \kappa(\theta_i,\ x) \cdot \ln \frac{\kappa(\theta_i,\ x)}{\kappa(\theta_{1-i},\ x)} \\ &\quad \sum_{i=0,1} \sum_{C \in \phi(\mathbf{c}_X)} \kappa(\{\theta_i\} \times C) \cdot \ln \frac{\kappa(\{\theta_i\} \times C)}{\kappa(\{\theta_{1-i}\} \times C)} \end{aligned} \tag{3.4.112}$$

From (3.4.52), the above is also restated as

$$\Delta_1(\mathbf{c}_X) = S_\Theta(X) \quad S_\Theta(X : \mathbf{c}_X), \tag{3.4.113}$$

where

$$\begin{aligned} S_\Theta(X) &\triangleq \sum_{x \in X} \{\ \lambda(\theta_0) \cdot \nu_X(x \mid \theta_0) \cdot \ln \frac{\nu_X(x \mid \theta_0)}{\nu_X(x \mid \theta_1)} \\ &\quad + \lambda(\theta_1) \cdot \nu_X(x \mid \theta_1) \cdot \ln \frac{\nu_X(x \mid \theta_1)}{\nu_X(x \mid \theta_0)}\ \} \\ &= \sum_{C \in \phi(\mathbf{1}_X)} \{\lambda(\theta_0) \cdot \nu_X(C \mid \theta_0) \cdot \ln \frac{\nu_X(C \mid \theta_0)}{\nu_X(C \mid \theta_1)} \end{aligned}$$

$$+ \lambda(\theta_1)\cdot\nu_X(C \mid \theta_1)\cdot \ln \frac{\nu_X(C \mid \theta_1)}{\nu_X(C \mid \theta_0)} \}, \tag{3.4.114}$$

$$S_\Theta(X : \mathbf{c}_X) = \sum_{C\in\phi(\mathbf{c}_X)} \{\lambda(\theta_0)\cdot\nu_X(C \mid \theta_0)\cdot \ln \frac{\nu_X(C \mid \theta_0)}{\nu_X(C \mid \theta_1)}$$

$$+ \lambda(\theta_1)\cdot\nu_X(C \mid \theta_1)\cdot \ln \frac{\nu_X(C \mid \theta_1)}{\nu_X(C \mid \theta_0)} \}. \tag{3.4.115}$$

Comparing with (3.2.16), $S_\Theta(X)$ is the mean information intensity of X with respect to discrimination between H_0: $\theta = \theta_0$ and H_1: $\theta \quad \theta_1$. Also, $S_\Theta(X : \mathbf{c}_X)$ can be interpreted as the mean information intensity of X under the compression (partition) $\mathbf{c}_X$. Therefore, the measure Δ_1 can be interpreted as the decrease of the mean information intensity of the input information X due to the compression $\mathbf{c}_X$.

On the otherhand, Δ_2 can be rewritten as

$$\Delta_2(\mathbf{c}_X) = \sum_{i=0,1} \sum_{\substack{C\in\phi(\mathbf{c}_X) \\ D\in\phi(\mathbf{d}_Y[\mathbf{c}_X])}} \hat{\tau}(\{\theta_i\}\times C\times D\cap\Xi_a)\cdot \ln \frac{\hat{\tau}(\{\theta_i\}\times C\times D\cap\Xi_a)}{\hat{\tau}(\{\theta_{1-i}\}\times C\times D\cap\Xi_a)}$$

$$\sum_{i=0,1} \sum_{C\in\phi(\mathbf{c}_X)} \hat{\tau}(\{\theta_i\}\times C\times Y\cap\Xi_a)\cdot \ln \frac{\hat{\tau}(\{\theta_i\}\times C\times Y\cap\Xi_a)}{\hat{\tau}(\{\theta_{1-i}\}\times C\times Y\cap\Xi_a)}$$

$$= \delta_1(\mathbf{c}_X) + \delta_2(\mathbf{c}_X), \tag{3.4.116}$$

where

$$\delta_1(\mathbf{c}_X) \triangleq \sum_{i=0,1} \sum_{\substack{C\in\phi(\mathbf{c}_X) \\ D\in\phi(\mathbf{d}_Y[\mathbf{c}_X])}} \hat{\tau}(\{\theta_i\}\times C\times D\cap\Xi_a)\cdot \ln \frac{\hat{\tau}(\{\theta_i\}\times C\times D\cap\Xi_a)}{\hat{\tau}(\{\theta_i\}\times C\times Y\cap\Xi_a)}, \tag{3.4.117}$$

$$\delta_2(\mathbf{c}_X) \triangleq \sum_{i=0,1} \sum_{\substack{C\in\phi(\mathbf{c}_X) \\ D\in\phi(\mathbf{d}_Y[\mathbf{c}_X])}} \hat{\tau}(\{\theta_i\}\times C\times D\cap\Xi_a)\cdot \ln \frac{\hat{\tau}(\{\theta_{1-i}\}\times C\times Y\cap\Xi_a)}{\hat{\tau}(\{\theta_{1-i}\}\times C\times D\cap\Xi_a)}. \tag{3.4.118}$$

The measure δ_1 is related to the so-called chi-square test and evaluates the accuracy of $\hat{\nu}_X$ to ν_X(cf. Jeffreys[1946]). When the numbers of data n_0 and n_1 are large, then δ_1 can be approximated as follows(for details, refer to Kullback[1952, pp.109-116], Carlton[1969], Pfaffelhuber[1971], and Nemetz[1972]).

$$\delta_1(\mathbf{c}_X) \simeq \frac{1}{2}\cdot\left(\frac{\lambda(\theta_0)}{n_0} + \frac{\lambda(\theta_1)}{n_1}\right)\cdot(\#(\phi(\mathbf{c}_X)) \quad 1). \tag{3.4.119}$$

In the above, $\#(\phi(\mathbf{c}_X))$ means the number of the atoms in partition $\mathbf{c}_X$.

On the other hand, it is difficult to evaluate the measure δ_2 and only the following qualitative property is known, which is derived from the independece between Y_0 and Y_1 and also from Jensen inequality.

$$\delta_2(\mathbf{c}_X) \geq 0 \quad \text{for any } \mathbf{c}_X \in \Pi(X). \tag{3.4.120}$$

Therefore, we finally obtain the following measure $\tilde{d}$ which is an underestimate of d and is derived by substituting the right side of (3.4.119) for $\delta_1(\mathbf{c}_X)$ and 0 for $\delta_2(\mathbf{c}_X)$, i.e.,

$$\tilde{d}(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X], \Pi_o) = (S_\Theta(X) - S_\Theta(X : \mathbf{c}_X)) + \frac{1}{2}\cdot\left(\frac{\lambda(\theta_0)}{n_0} + \frac{\lambda(\theta_1)}{n_1}\right)\cdot$$

$$(\#(\phi(\mathbf{c}_X)) - 1). \tag{3.4.121}$$

The suboptimum compression(partition) $\mathbf{c}_X^{**}$ is given as partition $\mathbf{c}_X$ which minimizes the above $\tilde{d}$. The first term of $\tilde{d}$ is the aforementioned information loss, i.e., the decrease of the mean information intensity due to the compression $\mathbf{c}_X$, and the second term is related to the accuracy of the *a priori* information Y. If the numbers n_0 and n_1 of the *a priori* samples are large, then the coefficient $1/2\cdot(\lambda(\theta_0)/n_0 + \lambda(\theta_1)/n_1)$ is small, and the effect of the first term is relatively large, and as a result, the number of the categories become large. This relationship is closely related to Kanal's result(1968). In any case, the suboptimum category aggregation is determined by the mean information intensity $S_R(S_\Theta)$ and the number of the categories $\#(\phi(\mathbf{c}_X))$.

To clarify the properties of the suboptimum category aggregation method, we next consider the case where the categories in an item are represented as a continuum. In this case, the aggregation problem is regarded as quantization problem of the continuum.

3.4.2.3 Optimum Quantization Method for the Continuum of Categories

In this section, we consider the optimum aggregation of the categories in an item in which the categories constitute a continuum. This is a kind of optimum quantization problem and the similar problems were discussed by Max (1960), Gish & Pierce(1968), and Wood(1969) in the framework of the rate-distor-

tion theory. However, our approach is different from theirs in the sense that their approach is focused upon the certain distance between the quantized data and the original data, while our approach is based on the discrimination information as discussed in the previous sections. The discussions in this section clarify the properties of the optimum aggregation method of categories and particularly the relation of the optimum aggregation with the quantification ρ introduced in Section 3.2(cf. Katai, Imanaga, & Iwai[1972b & 1974]).

Suppose that the set X of the categories in an item X is given by an interval, say, [a, b] in the real axis. The probability measures $\nu_X(x \mid R)$ and $\nu_X(x \mid \bar{R})$(which corresponds to $\nu_X(x \mid \theta_0)$ and $\nu_X(x \mid \theta_1)$, respectively in Section 3.4.2.1) are represented as the probability distribution functions as follows:

$$p_0(x) \triangleq \frac{d\nu_X(x \mid R)}{dx},$$

$$p_1(x) \triangleq \frac{d\nu_X(x \mid \bar{R})}{dx}, \qquad \text{for } x \in [a, b] \tag{3.4.122}$$

In the above, we suppose that $\nu_X(\cdot \mid R)$, $\nu_X(\cdot \mid \bar{R})$, and the Lebesgue measure on [a, b] are absolutely continuous to each other.

Next, we assume that X is partitioned(quantized) into (finite) M intervals by the points $a = x_0 < x_1 < x_2 < \cdots < x_{M-1} < x_M = b$, i.e.,

$$\phi(\mathbf{c}_X) = \{[a, x_1], [x_1, x_2], \cdots, [x_{M-1}, b]\}. \tag{3.4.123}$$

To calculate $L(x : \mathbf{c}_X)$, we introduce the following histogram of p_i for i 0, 1.

$$\begin{aligned}\bar{p}_i(x) &\triangleq \nu_X(F_j \mid \theta_i) / |F_j| \\ &= \int_{x_{j-1}}^{x_j} p_i(x)\, dm_j(x) \\ &= E_{m_j}[p_i] \quad \text{for i } 0, 1 \text{ and } x \in F_j \triangleq [x_{j-1}, x_j],\end{aligned} \tag{3.4.124}$$

where

$$dm_j(x) \triangleq \frac{dx}{|F_j|}, \tag{3.4.125}$$

$$|F_j| \triangleq x_j - x_{j-1}. \tag{3.4.126}$$

Using the above histograms, $L(x : \mathbf{c}_X)$ defined by (3.4.56) is given as

$$L(x : \mathbf{c}_X) \quad \frac{P(R)}{P(\bar{R})} \cdot \frac{\bar{p}_0(x)}{\bar{p}_1(x)}, \tag{3.4.127}$$

where $P(R)$ and $P(\bar{R})$ correspond to $\lambda(\theta_0)$ and $\lambda(\theta_1)$, respectively.

We denote by $\gamma = (\gamma_{01}, \gamma_{02}, \cdots, \gamma_{0n_0} : \gamma_{11}, \gamma_{12}, \cdots, \gamma_{1n_1})$ the quantized sample value of the *a priori* information Y, i.e.,

$$\gamma_{ij} \quad m \quad \text{iff} \quad x_{ij} \in F_m \quad \text{for } i = 0, 1, \text{ and } m = 1, 2, \cdots, M, \tag{3.4.128}$$

where x_{ij} is the sample value of X_{ij}(cf. (3.4.96) and (3.4.99)). In the usual cases, the sequences $X_{01}, X_{02}, \cdots, X_{0n_0}$ and $X_{11}, X_{12}, \cdots, X_{1n_1}$ are Bernoulli sequences, and γ have the following probability structure.

$$\mu_\gamma(\gamma) = \prod_{k=1}^{n_0} \nu_X(F_{\gamma_{0k}} \mid \theta_0) \cdot \prod_{k=1}^{n_1} \nu_X(F_{\gamma_{1k}} \mid \theta_1). \tag{3.4.129}$$

In relation to (3.4.124), we consider the histogram on X given by γ as follows:

$$\bar{\hat{p}}_i(x \mid \gamma) = \sum_{k=1}^{n_i} \frac{\chi_j(\gamma_{ik})}{n_i \cdot |F_j|} \quad \text{when } x \in F_j \text{ for } i = 0,1 \text{ and } j = 1, 2, \cdots, M, \tag{3.4.130}$$

where

$$\chi_j(\gamma_{ik}) \quad \begin{cases} 1 & \text{if } \gamma_{ik} = j \\ 0 & \text{if } \gamma_{ik} \neq j. \end{cases} \tag{3.4.131}$$

By the use of the above histograms, $\hat{L}(x : \mathbf{c}_X, y)$ defined by (3.4.63) is given as

$$\hat{L}(x : \mathbf{c}_X, y) = \frac{P(R)}{P(\bar{R})} \cdot \frac{\bar{\hat{p}}_0(x \mid \gamma)}{\bar{\hat{p}}_1(x \mid \gamma)}. \tag{3.4.132}$$

The unbiasedness (3.4.102) corresponds to the next averaging property(cf. Cover[1971]).

$$E_{\mu_\gamma}[\bar{\hat{p}}_i(x \mid \gamma)] = \bar{p}_i(x) \quad \text{for } i = 0, 1. \tag{3.4.133}$$

In the above discussion, we have introduced three probability distribution functions $p_i(x)$, $\bar{p}_i(x)$, and $\bar{\hat{p}}_i(x \mid \gamma)(i = 0, 1)$ on $X(= [a, b])$. It is easy to see that p_i, $\bar{p}_i$, and $\hat{p}_i$ correspond to the partitions $\mathbf{1}_X \times \mathbf{0}_Y$, $\mathbf{c}_X \times \mathbf{0}_Y$, and $\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X]$, respectively. The averaging properties (3.4.124) and (3.4.132)

join the three probabilities. Fig. 3.4.5 depicts the above relationships.

The above probabilities give three different quantifications for X as follows:

$$\rho(x) \triangleq \frac{p_0(x)}{p_1(x)}, \tag{3.4.134}$$

$$\bar{\rho}(x) \triangleq \frac{\bar{p}_0(x)}{\bar{p}_1(x)}, \tag{3.4.135}$$

$$\bar{\hat{\rho}}(x \mid \gamma) \triangleq \frac{\bar{\hat{p}}_0(x \mid \gamma)}{\bar{\hat{p}}_1(x \mid \gamma)}. \tag{3.4.136}$$

The quantification $\bar{\hat{\rho}}(x \mid \gamma)$ corresponds to the actual quantification for the category x under the quantization (3.4.123) and under the *a priori* information γ, while $\rho(x)$ corresponds to the "ideal" quantification for x provided that the true probabilities p_0 and p_1 are known. Also, $\bar{\rho}(x)$ corresponds to the ideal quantification for x under the quantization $\mathbf{c}_X$.

The difference measures $\Delta_1(\mathbf{c}_X)$ $\Delta(\mathbf{1}_X \times \mathbf{0}_Y, \mathbf{c}_X \times \mathbf{0}_Y)$ and $\Delta_2(\mathbf{c}_X)$ $\Delta(\mathbf{c}_X \times \mathbf{d}_Y[\mathbf{c}_X], \mathbf{c}_X \times \mathbf{0}_Y)$ are calculated as follows:

$$\Delta_1(\mathbf{c}_X) = \int_a^b (P(R)p_0(x) \quad P(\bar{R})p_1(x))(\rho(x) \quad \bar{\rho}(x))\, dx, \tag{3.4.137}$$

$$\Delta_2(\mathbf{c}_X) = E_{\mu_Y}\Big[\int_a^b (P(R)p_0(x \mid \gamma) - P(\bar{R})p_1(x \mid \gamma)) \cdot (\bar{\hat{\rho}}(x \mid \gamma) \quad \bar{\rho}(x))\, dx \Big]. \tag{3.4.138}$$

That is to say, these two measures evaluate the differences between ρ and $\bar{\rho}$, and between $\bar{\hat{\rho}}$ and $\bar{\rho}$, respectively. The measure defined by (3.4.117) is, in this case, calculated as below.

$$\delta_1(\mathbf{c}_X) = E_{\mu_Y}\Big[\int_a^b \{P(R)\bar{\hat{p}}_0(x \mid \gamma) \cdot \ln \frac{\bar{\hat{p}}_0(x \mid \gamma)}{\bar{p}_0(x)} + P(\bar{R})\bar{\hat{p}}_1(x \mid \gamma) \cdot \ln \frac{\bar{\hat{p}}_1(x \mid \gamma)}{\bar{p}_1(x)} \} \, dx \Big]. \tag{3.4.139}$$

By the use of the following approximation

$$\text{if } \alpha \simeq \beta, \text{ then } \ln \frac{\alpha}{\beta} \simeq \frac{\alpha^2 \quad \beta^2}{2\alpha\beta}, \tag{3.4.140}$$

we obtain

$$\delta_1(\mathbf{c}_X) \simeq E_{\mu_Y}[P(R) \int_a^b \frac{(\hat{\bar{p}}_0(x \mid \gamma) \quad \bar{p}_0(x))^2}{2\bar{p}_0(x)} dx$$

$$+ P(\bar{R}) \int_a^b \frac{(\hat{\bar{p}}_1(x \mid \gamma) \quad \bar{p}_1(x))^2}{2\bar{p}_1(x)} dx]. \tag{3.4.141}$$

In the above,

$$\int_a^b \frac{(\hat{\bar{p}}_0(x \mid \gamma) \quad \bar{p}_0(x))^2}{\bar{p}_0(x)} dx \quad \text{and} \quad \int_a^b \frac{(\hat{\bar{p}}_1(x \mid \gamma) \quad \bar{p}_1(x))^2}{\bar{p}_1(x)} dx \tag{3.4.142}$$

are the χ^2-statistics with freedom M 1 denoted by χ^2_{M-1}, which evaluate the differences between the empirical distribution $\hat{\bar{p}}_i(x \mid \gamma)$ and the true distribution $\bar{p}_i(x)$ for i = 0, 1.

It is wellknown that

$$E[\chi^2_{M-1}] = \frac{M \quad 1}{n}, \tag{3.4.143}$$

where n is the number of data. Therefore, we finally obtain

$$\delta_1(\mathbf{c}_X) \sim \frac{1}{2} \cdot \left(\frac{P(R)}{n_0} + \frac{P(\bar{R})}{n_1} \right)(M \quad 1). \tag{3.4.144}$$

From (3.4.121), (3.4.137) and (3.4.144), the optimum quantization is given by minimizing

$$\tilde{d} \quad (S_R(X) \quad S_R(X : \mathbf{c}_X)) + \frac{1}{2} \cdot \left(\frac{P(R)}{n_0} + \frac{P(\bar{R})}{n_1} \right)(M \quad 1)$$

$$= \int_a^b (P(R)p_0(x) \quad P(\bar{R})p_1(x))(\rho(x) \quad \bar{\rho}(x)) \, dx$$

$$+ \frac{1}{2} \cdot \left(\frac{P(R)}{n_0} + \frac{P(\bar{R})}{n_1} \right)(M \quad 1) \tag{3.4.145}$$

with respect to M, x_1, x_2, $\cdots$, and x_{M-1}.

The second term of the right side of (3.4.145) depends only on the number of quantized levels M. Hence, if the number M is fixed beforehand, then the optimum quantizing points x_1, x_2, $\cdots$, x_{M-1} are given by minimizing the first term $S_R(X) \quad S_R(X : \mathbf{c}_X)$ in (3.4.145). For instance, let us consider the bio-

chemical data mentioned in the last part of Section 3.2. They are composed of 12 items, and we consider the optimum quantizing points for each item X_i (i = 1, 2, ···, 12). Let the number of quantized levels for X_i be 3 (which is the same as the usual quantization into i_1[less than normal], i_2[normal], and i_3[more than normal] as aforementioned) for i = 1, 2, ···, 12. In this case, the optimum quantizing points x_1 and x_2 in X_i are determined for each disease R_j by minimizing $S_{R_j}(X_i)$ $S_{R_j}(X_i : \mathbf{c}_{X_i})$ i.e., by maximizing $S_{R_j}(X_i : \mathbf{c}_{X_i})$, for i = 1, 2, ···, 12 and j 1, 2, ···, 5. Fig. 3.4.6 shows the resultant quantizing points for j - 2 (renal failure). The encircled area in each item is the interval $[x_1, x_2]$. Also, Fig. 3.4.7 shows the values of $S_{R_j}(X_i : \mathbf{c}_{X_i})$ for j = 2 and i = 5 (uric acid) versus $\mathbf{c}_{X_i}$, i.e., the values of x_1 and x_2. In this case, $S_{R_j}(X_i : \mathbf{c}_{X_i})$ has quite complex characteristics and has multiple peaks. By the use of the above optimum quantization, the Bayesian discrimination rate is, on the average (for R_1, R_2, ···, R_5), 73 %, while the usual quantization method yields 63 %. Namely, we can expect a 10% increase of the discrimination rate due to the optimum quantization (for further details, refer to Takagi, Katai, Iwai et al.[1972]). Moreover, we compared the discrimination rates by the optimum quantization into 3 and 5 levels in relation to the number of data (*a priori* information). It was observed that the 3-level quantization method yields a higher discrimination rate than the 5-level method when the number of data is less than 100. In the other case, the 5-level method is better than the 3-level method (for details, refer to Katai, Iwai, et al.[1973]).

In general, the algorithm to search for the optimum values of M, x_1, x_2, ···, and x_{M-1} consists of two steps. In the first step, the values of x_1, x_2, ···, and x_{M-1} are seeked to attain the minimum value of $S_R(X)$ $S_R(X : \mathbf{c}_X)$ for various positive numbers M's. In the second step, we calculate the sum of the minimum value (obtained at the first step) and the value $1/2 \cdot (P(R)/n_0 + P(\overline{R})/n_1) \cdot$ (M 1) for each value of M. The optimum quantization is given by the value of M, x_1, x_2, ···, and x_{M-1} which attain the minimum of the above sum.

In the sequel, we examine the properties of the above optimum quantization method by using an example which is similar to the example in Section 3.3.2 (cf. Katai, Imanaga, & Iwai[1974]). In this example, we consider k^2 $k \times k$ items each of which has the following truncated normal distributions. Namely, the item X_{ij} (i = 1, 2, ···, k and j 1, 2, ···, k) is given by truncating a random variable by a and b ($a < b$), which has normal distribution $N(\mu, \sigma)$ in the sample type R and $N(\mu_{ij}, \sigma_{ij})$ in the sample type $\overline{R}$ (cf. Fig. 3.4.8). In the above, the mean μ_{ij} and the standard deviation σ_{ij} are set as

$$\mu_{ij} = \frac{i \quad 1}{k \quad 1} \cdot RAM + \mu,$$

$$\sigma_{ij} = \frac{j \quad 1}{k \quad 1} \cdot RAV + \sigma \quad \text{for } i, j = 1, 2, \cdots, k. \tag{3.4.146}$$

Hence each item has different probabilistic characteristics which can be altered by changing RAM and RAV.

First, we consider the optimum quantization of all the items subject to given M(the number of quantizing levels). The value of the parameters are set as P(R) = P($\bar{R}$) 1/2, k = 4, a = 5.0, b = 5.0, μ= 0, σ = 1.0, and RAM = RAV = 0.9(to reduce the effect of truncation, we set the values of σ, RAM, and RAV as small compared to |a| and |b|). We compared the optimum quantizations with the uniform quantizations. Fig. 3.4.9 shows the results for item X_{43} and for M = 2, 3, and 4.

It will be readily understood that the optimum quantization points(indicated by open circles) are distributed to the region where the curve of the quantification ρ is steep. This property will be examined in the latter part of this section. We also compared the optimum and the uniform quantizations in terms of the discrimination rate P_c and the iformation loss $\Delta_1 = S_R(\mathbf{X}) \quad S_R(\mathbf{X} : \mathbf{c}_X) \quad \sum_{i,j=1}^{4} S_R(X_{ij}) \quad S_R(X_{ij} : \mathbf{c}_{X_{ij}})$(, where $\mathbf{X} - (X_{11}, X_{12}, \cdots, X_{44})$). Table 3.4.2 shows the resultant values for M = 2, 3, and 4. In the table, the discrimination rate is estimated by using 1,000 test inputs for each type R or $\bar{R}$. It is obvious that there exists a profound difference between the optimum and the uniform quantizations with respect to both P_c and Δ_1.

In the above, we considered the optimun quantization problems subject to given number M of quantizing levels based on the measure Δ_1. Next we consider the optimum number M of quantizing levels subject to the uniform quantization method based on the criterion $\tilde{d}(\Delta_1 + \delta_1)$ given by (3.4.145). To obtain accurate estimated value of P_c and to assure the validity of the approximate formula (3.4.144), we set the values of the parameters as P(R) = P($\bar{R}$) 1/2, k = 3, a -3.0, b = 3.0, μ = 0, σ = 3.0, RAM RAV 1.0, n_0 n_1 5,000.(The alteration of k from 4 to 3 is for assuring the accuracy of P_c. To assure the validity of (3.4.144), the number of samples in each quantized interval must be sufficiently large, hence we set n_0 n_1 - 5,000, the values of |a| and |b| being decreased and the value of σ increased.) We consider the uniform quantizations for M = 2 to 128 levels. Fig. 3.4.10 shows the relationship between the measure $\tilde{d}$ and the discrimination rate P_c.

Roughly speaking, the quatization level M with minimum $\tilde{d}$(information loss) attains the maximum discrimination rate for each case. In the figure, D(M) is an approximatation of $\tilde{d}$ as introduced in the sequel.

In the following, we consider the asymptotic properties of the information loss $\tilde{d}$ for the uniform quantization when the number of quantizing level M is sufficiently large(cf. Katai, Imanaga, & Iwai[1974]).

By taking note of (3.4.124) and by using Taylor expansion, we have

$$\int_{x_j}^{x_{j+1}} (P(R)p_0(x) - P(\bar{R})p_1(x))(\rho(x) \quad \bar{\rho}(x))\,dx$$
$$- \frac{h^3}{24} p(x_j)\rho'(x_j)^2 + O(h^4), \tag{3.4.147}$$

where h is the length of the quantized interval, i.e., h (b a)/M, and p(x) and $\rho'(x)$ are given by

$$p(x) \triangleq P(R)p_0(x) + P(\bar{R})p_1(x), \tag{3.4.148}$$

$$\rho'(x) \triangleq \frac{d\rho(x)}{dx} = \frac{p_0'(x)}{p_0(x)} \quad \frac{p_1'(x)}{p_1(x)}. \tag{4.3.149}$$

From (3.4.137), we obtain

$$\Delta_1(\mathbf{c}_X) \simeq \Delta_M, \tag{3.4.150}$$

$$\Delta_M \triangleq \frac{(b \quad a)^2}{12} \frac{I_F}{M^2}, \tag{3.4.151}$$

where I_F is defined as

$$I_F \triangleq \frac{1}{2} \cdot \int_a^b p(x)\cdot\rho'(x)^2\,dx = \frac{1}{2}\cdot E[\rho'(x)^2] \tag{3.4.152}$$

and can be considered as an extension of Fisher's information for the translation parameter case and coincides with the Fisher's information of p_1 when p_2 is the uniform distribution in [a, b] and P(R) tends to 1(for details, refer to Fisher[1959] and Shepp[1965]).

In the following, we assume that the relative frequencies of the sample types R and $\bar{R}$ in the *a priori* information Y coincide with the *a priori* probabilities of R and $\bar{R}$, i.e.,

$$\frac{n_0}{n_0 + n_1} = P(R), \quad \frac{n_1}{n_0 + n_1} = P(\bar{R}). \tag{3.4.153}$$

In this case, we have

$$\delta_1(\mathbf{c}_X) \simeq \frac{M \quad 1}{N}, \tag{3.4.154}$$

where N is the total number of the samples in the *a priori* information, i.e.,

$$N = n_0 + n_1. \tag{3.4.155}$$

From (3.4.145), the measure $\tilde{d}$ can be approximated by the following quantity D(M) for sufficiently large M.

$$\begin{aligned} D(M) &\triangleq \Delta_M + \frac{M \quad 1}{N} \\ &= \frac{(b \quad a)^2}{12} \cdot \frac{I_F}{M^2} + \frac{M \quad 1}{N}. \end{aligned} \tag{3.4.156}$$

As shown in Fig. 3.4.10, D(M) almost coincides with the true value $\tilde{d}$ for $M \geq 8$, and the approximate fomulae (3.4.150) and (3.4.156) seem to be precise approximations.

In the above, we examined the information loss by the uniform quantization when the number M of quantizing levels is sufficiently large. Next, we consider the information loss of the optimum quantization when M tends to infinity. Namely, we consider the following idealized quantization scheme , where the quantization points are given by a distribution function. Let T(x) be the distribution frnction, i.e., the number of quantization points in the interval $(x \quad dx, x + dx)(\subset [a, b])$ be $T(x) \cdot 2dx$. Then, from (3.4.147), the information loss $\Delta(T)(\triangleq \Delta_1(\mathbf{c}_X))$ is given by

$$\Delta(T) \quad \frac{1}{24} \cdot \int_a^b p(x) \left(\frac{\rho'(x)}{T(x)}\right)^2 dx. \tag{3.4.157}$$

We call the quantization, which attains the minimum value of $\Delta(T)$ under the condition that the number of quantization points(quantizing levels) is M, the *asymptotically optimum quantization* with number of levels M. The above condition corresponds to

$$\int_a^b T(x)\, dx = M. \tag{3.4.158}$$

This is given by the calculus of variations, and the solution T*(x) is given as

$$T^*(x) = M \cdot \left(\int_a^b z(x)\, dx \right)^{-1} \cdot z(x), \tag{3.4.159}$$

where

$$z(x) \triangleq (p(x)\cdot\rho'(x)^2)^{1/3}. \tag{3.4.160}$$

From the definition, it is obvious that the asymptotically optimum quantization asymptotically coincides with the suboptimum quantization discussed before when the number of quantizing levels M tends to infinity. Also, from (3.4.159) and (3.4.160), it is easy to see that the asymptotically optimum quantization can be regarded as the *minimum distortion quantization*, i.e., the quantization which yields the minimum value of the mean square difference between the original value of x and the quantized value of x(which is given as the middle point of the interval to which x belongs) under the following metric: The metric(distance) ds between x and x + dx is given as

$$ds \triangleq |\rho'(x)|\cdot dx = |d\rho(x)|. \tag{3.4.161}$$

That is, the quantification ρ plays the essential role in the suboptimum quantizations, and hence in the suboptimum category aggregation.

In the following, we transform the interval [a, b] into [1, 0] by the affine transformation in order to make the mathematical expressions simple, for linear transformations have no effect on the quantization problems. Let Δ_M^* denote the information loss by the asymptotically optimum quantization with M levels. Then, from (3.4.157) and (3.4.159), Δ_M^* is approximated as

$$\Delta_M^* \simeq \frac{1}{24}\cdot\frac{1}{M^2}\cdot\left(\int_0^1 z(x)\,dx\right)^3. \tag{3.4.162}$$

The corresponding approximation of $\tilde{d}$ is given as

$$\begin{aligned} D^*(M) &\triangleq \Delta_M^* + \frac{M}{N}\frac{1}{} \\ &= \frac{1}{24}\cdot\frac{1}{M^2}\cdot\left(\int_0^1 z(x)\,dx\right)^3 + \frac{M}{N}\frac{1}{}. \end{aligned} \tag{3.4.163}$$

Let us compare the efficiencies of the suboptimum and the uniform quantizations subject to fixed M by their information losses as follows:

$$\frac{\Delta_M^*}{\Delta_M} \quad \eta^3, \tag{3.4.164}$$

where

$$\eta \triangleq \frac{\|z\|_1}{\|z\|_3}, \tag{3.4.165}$$

$$\| z \|_p \triangleq \left(\int_0^1 z(x)^p \, dx \right)^{1/p} \tag{3.4.166}$$

From the Hölder inequality, we have

$$\eta \leq 1. \tag{3.4.167}$$

Hence the suboptimum quantization is more efficient than the uniform quantization, and its degree of dominance is determined by $z(x)(0 \leq x \leq 1)$.

Next, we consider the optimum number of quantizing levels for the uniform and the asymptotically optimum quantizations. Let them be denoted by M_{opt} and M^*_{opt}, respectively. Then, from (3.4.156) and (3.4.163), they are approxiamted by

$$\begin{aligned} M_{opt} &\quad (12)^{-1/3} \cdot N^{1/3} \cdot \| z \|_3 , \\ M^*_{opt} &= (12)^{-1/3} \cdot N^{1/3} \cdot \| z \|_1 . \end{aligned} \tag{3.4.168}$$

In both cases, the optimum M is proportional to $N^{1/3}$, the cubic root of the number of samples in the *a priori* information(experiment). Hence we need eight times the number of samples to make the optimum number of quantizing levels twice.

When the asymptotically optimum and the uniform quantizations are compared with respect to the values of $\tilde{d}$ at their optimum quantizing levels, then the comparison is

$$\frac{D^*(M^*_{opt})}{D(M_{opt})} = \frac{M^{*\,2/3}_{opt}}{M^{2/3}_{opt}} \simeq \frac{M^*_{opt}}{M_{opt}} = \eta \leq 1. \tag{3.4.169}$$

Namely, the comparison in this case is also prescribed by the parameter η defined by (3.4.165), and the dominance of the asymptotically optimum quantization to the uniform quantization is weaker compared to the domiance indicated by (3.4.164)

In Section 3.4, we have discussed efficient ways of aggregation of items and category types in response matrices. Briefly speaking, these methods can be prescribed by certain measures which evaluate the mean difference beween the original quantifications and the quantifications based on the aggregation for items or category types. For item aggregation problems, the correlation measure C_R in Section 3.4.1 corresponds to the above measure, and for category aggregation problems, the measure of information loss $\Delta_1(\mathbf{c}_\chi)$ or $\tilde{d}$ corresponds to it. Moreover, we considered the category aggregation methods in the more gen-

eral setting of the problem, i.e., by taking note of the accuracy of the response matrices due to the deficiency of *a priori* information. This setting of the problem also has relevance to the item aggregation problems. However, there is no general framework such as the lattice structure of information(as in the case of category aggregation problems) for the item aggregation problems.

3.5 Application of the Quantification Method to the Analyses of Luce's Learning Model

In Section 3.2, we introduced the quantification $\rho(X_i)$ or $\rho(\mathbf{X})$ to measure the effect of each item X_i or a collection of items $\mathbf{X}$ $(X_1, X_2, \ldots, X_n)$ on the discrimination between sample types R and $\bar{R}$. Also, we introduced the measures $I(R|\bar{R}\ ;\ \tilde{P}(X_i))$ or $I(R|\bar{R}\ ;\ \tilde{P}(\mathbf{X}))$ to evaluate the degree of belongingness of the population $\{X_i\}$ or $\{\mathbf{X}\}$ having probability laws $\tilde{P}(X_i)$ or $\tilde{P}(\mathbf{X})$ to R compared to $\bar{R}$. In Section 3.3, we discussed the discrimination rates of the statistical decision methods directly related to ρ such as likelihood ratio tests. In Section 3.4, the methods to determine the aggregation of items or categories are discussed. These studies are in a sense static; the items are given beforehand.

Let us consider the case where the items $\{X_i\}$ are given as a time series and take note of the changes of the values of the quantification ρ. This case is, in the general framework, considered to be a kind of discrimination learning process(cf. Bush & Mosteller[1955], Luce, Bush & Galanter[1965], Hilgard & Bower[1966] and Tsetlin[1973]).

The most fundamental constituents of learning processes are the *learning subjects* and the *environments*, and these two constituents interact with each other. The effect of the learning subjects on the environments is called *action* and the effect of the environments on the learning subjects is called *reinforcement*. Let the possible reinforcements by an environment be g_1, g_2, ... and g_k and also the actions permitted to a learning subject be f_1, f_2, ... and f_m. The case which is most often examined is the next situation composed of two actions f_1 and f_2, and two reinforcements g_1 (penalty) and g_2(non penalty). The learning subject learns to evade penalty g_1; it tends to choose one of actions f_1 or f_2 that causes less frequency of the penalty.

In this section, we propose a stochastic learning process model based on the following considerations. The learning subjects decide which action, f_1 or f_2, is preferable based on past experience, or in other words, which action is less penalized by the environment.

As the foundation of the decision of the preferability between f_1 and f_2, we suppose that the learning subject sets up two hypothetical situations R and $\bar{R}$ of the environment. Under situation R, f_1 is preferable (i.e., less penalized) to f_2. On the contrary, f_2 is preferable to f_1 under sutuation $\bar{R}$. The learning subject seeks which of the situations R or $\bar{R}$ fits the actual situation better than the other.

To embody the above consideration, R and $\bar{R}$ are set to be alternative hypotheses which could be held by the learning subject giving the *subjective probability structure* for the input-output relation of the random environment. That is to say, the learning subject sets up a response table (response matrix) composed of two items corresponding to actions f_1 and f_2,and of two categories (category types) corresponding to reinforcements g_1 and g_2, as shown in Table 3.5.1

In the table, item f_i corresponds to action f_i for i = 1,2, and category i_j corresponds to the occurrence of the reinforcement g_j caused by action f_i for i =1,2 and j = 1,2 . The probability $P(i_j|R)$ (or $P(i_j|\bar{R})$) is interpreted as the subjective conditional probability $\hat{P}(g_j|f_i, R)$ (or $\hat{P}(g_j|f_i,\bar{R})$) (held by the learning subject) of output g_j conditioned upon input f_i and hypothesis (sample type) R(or $\bar{R}$). Let X_n denote the action-reinforcement pair (random variable) taken by the learning subject at time n. Then the corresponding response table at time n is given by Table 3.5.2.

In the table, each item X_t (t 1,2, ..., n) has four categories x_{ij} - (f_i, g_j) (i - 1,2, j 1,2), and $P_t(f_i)$ (t = 1,2, ..., n, i = 1,2) represents the probability of action f_i (taken by the learning subject) at time t, which is determined by the *a posteriori* probabilities of the sample types R and $\bar{R}$ at time t-1.

In Section 3.5.1, under certain assumptions on the determination of the above *a posteriori* probabilities, we show that the rule of the renewal of the action probability $P_n(f_i)$ is the same as the so-called *beta model* of Luce. In Section 3.5.2, we discuss the learning behavior of the model by the use of the quantification ρ and the quantified mean I introduced in Section 3.2. Also, using the above measures, we examine the asymptotic behavior of learning process of the model. Particularly, the effect of the setting of the subjective conditional probabilities on the asymptotic behavior is examined in connection with actual conditional probabilities of the environment, referring to an excellent result by Norman(1970). As a result, a condition on the values of subjective conditional probabilities is derived in order to assure expedient behavior of the model. When the learning process is viewed as a Markov

chain with infinitely many states, the chain has interesting properties tightly related to Norman's result. Moreover, it has a tight connection with an interesting learning model which is called *linear tactics* and is derived by the Russian scientist Tsetlin(1961). A kind of random processes called *martingale processes*, introduced by Doob(1953), seems to have a desirable property when viewed as a stochastic learning process. We clarify the condition under which the behavior of the model constitutes a martingale process. In Section 3.5.3, using the measure I, the relationship between the complexity of a random environment and the memory capacity of a learning subject is examined with respect to its efficiency. In Section 3.5.4, we extend the model in such a way that the learning subject has more than two hypotheses or a *continuum of hypotheses*. The learning behaviors of the extended models are compared with that of the model with just two hypotheses.

3.5.1 Construction of a Conditional Probability Learning Model and Considerations on Its Equivalence with Luce's Beta Model

Fig, 3.5.1 shows the schematic diagram of our proposed stochastic learning model. In the figure, F_n denotes the random variable representing the selected action (the input to the environment) at the time n and G_n denotes the reinforcement (the corresponding output of the environment). The stochastic property of the random environment is represented by the (objective) probability law $P(G_n/F_n)$. We restrict ourselves to the case,

$$F_n \quad f_1 \text{ or } f_2 \ , \tag{3.5.1}$$

$$G_n \quad g_1 \text{ (penalty) or } g_2 \text{ (non-penalty)} \ . \tag{3.5.2}$$

The learning subject sets up two hypotheses (sample types) R and $\overline{R}$ as mentioned previously. These hypotheses determine the next action. The learning subject has subjective probabilities $\hat{P}(g_j/f_i, R)$ and $\hat{P}(g_j/f_i, \overline{R})$, where $\hat{P}(g_j/f_i, R)$ (or $\hat{P}(g_j/f_i, \overline{R})$) represents the subjective probability of output g_j under hypothesis R (or $\overline{R}$) and the input is f_i.

As aforementioned, the action-reinforcement pair at time t, i.e., the t^{th} item in the response table shown by Table 3.5.2, is denoted by X_t :

$$X_t \quad (F_t, G_t). \tag{3.5.3}$$

The progress of learning depends on the past action-reinforcement pairs $\mathbf{X}_n$ (sequence of items) :

$$\mathbf{X}_n \quad (X_1, X_2, \ \ldots \ , X_n). \tag{3.5.4}$$

The subjective *a posteriori* probabilities $\hat{P}(R/\mathbf{X}_{n-1})$ and $\hat{P}(\bar{R}/\mathbf{X}_{n-1})$ of the hypotheses R and $\bar{R}$ at time t = n-1, conditioned upon past experience $\mathbf{X}_{n-1}$, are renewed to $\hat{P}(R/\mathbf{X}_n)$ and $\hat{P}(\bar{R}/\mathbf{X}_n)$ based on X_n, the action-reinforcement pair at t n. The *a priori* subjective probabilities, $\hat{P}(R)$ and $\hat{P}(\bar{R})$, of R and $\bar{R}$ are appropriately assumed beforehand in accordance with *a priori* knowledge of the learning subject about the properties of the environment. When $X_n = (f_i, g_j)$, the revised (renewed) probability $\hat{P}(R/\mathbf{X}_n)$ ($\hat{P}(\bar{R}/\mathbf{X}_n)$) is determined by the preceding values $\hat{P}(R/\mathbf{X}_{n-1})$, $\hat{P}(\bar{R}/\mathbf{X}_{n-1})$ and the aforementioned subjective probabilities for the random environment $\hat{P}(g_j/f_i, R)$ and $\hat{P}(g_j/f_i, \bar{R})$.

In our model, the decision of the next action F_{n+1} is made by a random experiment at t = n+1 such as coin tossing with the probability of heads f_1 being equal to $\hat{P}(R/\mathbf{X}_n)$ and that of tails f_2 being equal to $\hat{P}(\bar{R}/\mathbf{X}_n)$, i.e., the objective probability law P_{n+1} of the next action F_{n+1} is

$$P_{n+1}(f_1) \quad P(F_{n+1} = f_1/\mathbf{X}_n) = \hat{P}(R/\mathbf{X}_n), \tag{3.5.5}$$

$$P_{n+1}(f_2) = P(F_{n+1} = f_2/\mathbf{X}_n) \quad \hat{P}(\bar{R}/\mathbf{X}n). \tag{3.5.5'}$$

For n 0,

$$P_1(f_1) = \hat{P}(R), \tag{3.5.6}$$

$$P_1(f_2) = \hat{P}(\bar{R}). \tag{3.5.6'}$$

The random experiment at t n is prescribed only by the past expericnce $\mathbf{X}_{n-1}$ and its result is independent of whether hypothesis R or $\bar{R}$ is presupposed by the learning subject. Therefore, the subjective probabilities of $\{F_n = f_1\}$ conditioned upon R and $\bar{R}$ are set to be equal to the objective probability given by (3.5.5), i.e.,

$$\left.\begin{aligned} \hat{P}(F_n = f_1/R, \mathbf{X}_{n-1}) &= P(F_n = f_1/\mathbf{X}_{n-1}) \\ \hat{P}(F_n = f_1/\bar{R}, \mathbf{X}_{n-1}) &= P(F_n = f_1/\mathbf{X}_{n-1}) \end{aligned}\right\}, \tag{3.5.7}$$

$$\left.\begin{aligned} \hat{P}(F_n = f_2/R, \mathbf{X}_{n-1}) \quad P(F_n = f_2/\mathbf{X}_{n-1}) \\ \hat{P}(F_n = f_2/\bar{R}, \mathbf{X}_{n-1}) \quad P(F_n \quad f_2/\mathbf{X}_{n-1}) \end{aligned}\right\}. \tag{3.5.7'}$$

Conditions on the subjective probabilities

1) The following inequalities on the subjective probabilities must be set in accordance with the conditions on R and $\bar{R}$ stated in the introductory part.

$$\hat{P}(g_1/f_1, R) < \hat{P}(g_1/f_2, R), \tag{3.5.8}$$

(action f_1 is less penalized than action f_2 under hypothesis R)

$$\hat{P}(g_1/f_1, \bar{R}) > \hat{P}(g_1/f_2, \bar{R}). \tag{3.5.9}$$

(action f_1 is more penalized than action f_2 under hypothesis $\bar{R}$)

Or, from the relation : $P(g_1/f_1, R) = 1 \quad P(g_2/f_1, R)$, the above inequalities can be written as follows:

$$\hat{P}(g_2/f_1, R) > \hat{P}(g_2/f_2, R), \tag{3.5.8'}$$

$$\hat{P}(g_2/f_1, \bar{R}) < \hat{P}(g_2/f_2, \bar{R}). \tag{3.5.9'}$$

2) We make the supposition that the learning subject does not change the subjective probability according to the time and the past experience, i.e., the subjective probability $\hat{P}(G_n \quad g_j/F_n \quad f_i, R, \mathbf{X}_{n-1})$ of the event $\{G_n = g_j\}$ provided that $\{F_n = f_i\}$ and the past experience $\mathbf{X}_{n-1}$ and under hypothesis R is independent of $\mathbf{X}_{n-1}$ for any n, i 1,2, and j = 1,2;

$$\hat{P}(G_n = g_j/F_n = f_i, R, \mathbf{X}_{n-1}) = \hat{P}(g_j/f_i, R) \quad \text{for any } n,\ i = 1,2,\ j = 1,2 \tag{3.5.10}$$

$$\hat{P}(G_n \quad g_j/F_n \quad f_i, \bar{R}, \mathbf{X}_{n-1}) = \hat{P}(g_j/f_i, \bar{R}) \quad \text{for any } n,\ i \quad 1,2,\ j = 1,2 \tag{3.5.11}$$

This supposition means that the learning subject has no memory functions, or in other words, he assumes the stationarity of the random environment implicitly (cf. Winkler & Murphy(1973)). In Section 3.5.3, we consider the effect of memory functions on the learning process, where the above supposition is altered such that the left-hand sides of (3.5.10) and (3.5.11) depend on a subsequence of the past experience $\mathbf{X}_{n-1}$.

Revision of a posteriori probabilities of the hypotheses

From the above conditions, the *a posteriori* probabilities $\hat{P}(R/\mathbf{X}_{n-1})$ and $\hat{P}(\bar{R}/\mathbf{X}_{n-1})$ are revised by the following rule:

Suppose that the learning subject selected action f_i at time n and the output caused by f_i is g_j, i.e.,

$$F_n \quad f_i \text{ and } G_n = g_j. \tag{3.5.12}$$

Then, from the theorem of conditional probability, we have

$$\hat{P}(R/\mathbf{X}_n) \ (= \hat{P}(R/G_n = g_j, F_n = f_i, \mathbf{X}_{n-1}) \ )$$

$$= \frac{\hat{P}(R, G_n = g_j/F_n = f_j, \mathbf{X}_{n-1})}{\hat{P}(G_n = g_j/F_n = f_i, \mathbf{X}_{n-1})}$$

$$= \frac{\hat{P}(R,\ G_n = g_j/F_n = f_i,\ \mathbf{X}_{n-1})}{\hat{P}(R,\ G_n = g_j/F_n = f_i,\ \mathbf{X}_{n-1}) + \hat{P}(\bar{R},\ G_n = g_j/F_n = f_i,\ \mathbf{X}_{n-1})}. \tag{3.5.13}$$

Also, we have

$$\hat{P}(R,\ G_n = g_j/F_n = f_i,\ \mathbf{X}_{n-1}) = \hat{P}(G_n = g_j/F_n = f_i,\ R,\ \mathbf{X}_{n-1}) \cdot \hat{P}(R/F_n = f_i,\ \mathbf{X}_{n-1}). \tag{3.5.14}$$

From stationarity (3.5.10), we obtain

$$\hat{P}(R,\ G_n = g_j/F_n = f_i, \mathbf{X}_{n-1}) = \hat{P}(g_j/f_i,\ R) \cdot \hat{P}(R/F_n = f_i,\ \mathbf{X}_{n-1}). \tag{3.5.15}$$

Moreover, from (3.5.7), (3.5.7'), and using Bayes theorem, we have

$$\begin{aligned}
&\hat{P}(R/F_n = f_i,\ \mathbf{X}_{n-1}) \\
&= \frac{\hat{P}(F_n = f_i/R,\ \mathbf{X}_{n-1}) \cdot \hat{P}(R/\mathbf{X}_{n-1})}{\hat{P}(F_n = f_i/R,\ \mathbf{X}_{n-1}) \cdot P(R/\mathbf{X}_{n-1}) + \hat{P}(F_n = f_i/\bar{R},\ \mathbf{X}_{n-1}) \cdot \hat{P}(\bar{R}/\mathbf{X}_{n-1})} \\
&= \frac{P(F_n = f_i/\mathbf{X}_{n-1}) \cdot \hat{P}(R/\mathbf{X}_{n-1})}{P(F_n = f_i/\mathbf{X}_{n-1}) \cdot \hat{P}(R/\mathbf{X}_{n-1}) + P(F_n = f_i/\mathbf{X}_{n-1}) \cdot \hat{P}(\bar{R}/\mathbf{X}_{n-1})} \\
&= \frac{\hat{P}(R/\mathbf{X}_{n-1})}{\hat{P}(R/\mathbf{X}_{n-1}) + \hat{P}(\bar{R}/\mathbf{X}_{n-1})} - \hat{P}(R/\mathbf{X}_{n-1}),
\end{aligned} \tag{3.5.16}$$

it follows that

$$\hat{P}(R,\ G_n = g_j/F_n = f_i,\ \mathbf{X}_{n-1}) \quad \hat{P}(g_j/f_i,\ R) \cdot \hat{P}(R/\mathbf{X}_{n-1}). \tag{3.5.17}$$

The above equation and (3.5.13) lead to

$$\hat{P}(R/\mathbf{X}_n) \quad \frac{\hat{P}(g_j/f_i,\ R) \cdot \hat{P}(R/\mathbf{X}_{n-1})}{\hat{P}(g_j/f_i,\ R) \cdot \hat{P}(R/\mathbf{X}_{n-1}) + \hat{P}(g_j/f_i,\ \bar{R}) \cdot \hat{P}(\bar{R}/\mathbf{X}_{n-1})}$$

$$\hat{P}(\bar{R}/\mathbf{X}_n) = 1 - \hat{P}(R/\mathbf{X}_n)$$

$$= \frac{\hat{P}(g_j/f_i,\ \bar{R}) \cdot \hat{P}(\bar{R}/\mathbf{X}_{n-1})}{\hat{P}(g_j/f_i,\ R) \cdot \hat{P}(R/\mathbf{X}_{n-1}) + \hat{P}(g_j/f_i,\ \bar{R}) \cdot \hat{P}(\bar{R}/\mathbf{X}_{n-1})}$$

when $F_n = f_i$ and $G_n = g_j$ for $i = 1,2$ and $j = 1,2$. (3.5.18)

Equivalence of the model with Luce's beta model

From (3.5.5) and (3.5.18), the recursive relation between the response probability P_{n+1} of the next action F_{n+1} and P_n of F_n is given by the following equation.

$$P_{n+1}(f_1) \quad \frac{\hat{P}(g_j/f_i, R)\cdot P_n(f_1)}{\hat{P}(g_j/f_i, R)\cdot P_n(f_1) + \hat{P}(g_j/f_i, \bar{R})\cdot P_n(f_2)}$$

$$= \frac{\dfrac{\hat{P}(g_j/f_i, R)}{\hat{P}(g_j/f_i, \bar{R})}\cdot P_n(f_1)}{\dfrac{\hat{P}(g_j/f_i, R)}{\hat{P}(g_j/f_i, \bar{R})}\cdot P_n(f_1) + (1 - P_n(f_1))},$$

$$P_{n+1}(f_2) \quad 1 \quad P_{n+1}(f_1),$$

when $F_n = f_i$ and $G_n = g_j$, for $i \quad 1, 2$ and $j = 1, 2$. (3.5.19)

The above innovation formula for the probabilities of the actions is equivalent to the nonlinear recursive formula known as Luce's beta model that is derived from his choice axiom as follows. In his theory of individual choice behaviour, Luce introduced a notion called *response strength* to express the frequencies of the choice actions of individuals. Let the response strength of the action f_i ($i = 1$ or 2) at the time n be $U_i(n)$. Then, from the choice axiom,

$$P_n(f_k) = \frac{U_k(n)}{U_1(n) + U_2(n)} \qquad \text{for } k = 1, 2. \tag{3.5.20}$$

He supposes a linear recursive relation between $U_k(n)$ and $U_k(n+1)$ as follows.

$$U_k(n+1) \quad \beta_{ij}^k U_k(n)$$

when $F_n = f_i$ and $G_n = g_j$, for $i = 1,2$, $j = 1,2$, and $k \quad 1,2$ (3.5.21)

where,

$$\beta_{ij}^k > 0. \tag{3.5.22}$$

Therefore, the relation between $P_{n+1}(f_k)$ and $P_n(f_k)$ is expressed by the next equation, for $k = 1, 2$.

$$P_{n+1}(f_1) = \frac{U_1(n+1)}{U_1(n+1) + U_2(n+1)} = \frac{\beta_{ij}^1 U_1(n)}{\beta_{ij}^1 U_1(n) + \beta_{ij}^2 U_2(n)}$$

$$= \frac{\beta_{ij}\hat{P}_n(f_1)}{\beta_{ij}\hat{P}_n(f_1) + (1 - \hat{P}_n(f_1))},$$

$$P_{n+1}(f_2) = 1 - P_{n+1}(f_1),$$

when $F_n = f_i$ and $G_n = g_j$, for $i = 1, 2$ and $j = 1,2$, (3.5.23)

where the coefficient β_{ij} is

$$\beta_{ij} = \frac{\beta^1_{ij}}{\beta^2_{ij}} \tag{3.5.24}$$

Therefore, our learning model is equivalent to Luce's beta model. The coefficient β_{ij} in the beta model corresponds to the following ratio of the subjective probabilities in our model.

$$\beta_{ij} = \frac{\beta^1_{ij}}{\beta^2_{ij}} = \frac{\hat{P}(g_j/f_i, R)}{\hat{P}(g_j/f_i, \bar{R})} . \tag{3.5.25}$$

In Luce's beta model, the coefficient β_{ij} has no concrete meaning; it is merely an adjustable parameter which determines the degree of the reinforcement. The above relation (3.5.25) throws light on the concrete interpretation of the coefficient.

3.5.2 Evaluation of the Efficiency of the Learning Model and Investigation of Its Asymptotic Learning Behavior based on the Quantification Method

In this section, we examine the rate of learning based on the quantification ρ and the measure I introduced in Section 3.2.

From the supposition on the determination of response probabilities (cf. Eqs. (3.5.5) and (3.5.5')), the learning behavior (successive changes of response probability $P_n(f_i)$ $(i = 1,2)$) is represented by the progression of the *a posteriori* subjective probabilities $\hat{P}(R/\mathbf{X}_n)$ and $\hat{P}(\bar{R}/\mathbf{X}_n)$ To measure the effect of past experience $\mathbf{X}_n$ (sequence of the occurred categories) on the above subjective *a posteriori* probabilities, we take note of the quantification ρ introdeced in Section 3.2, i.e.,

$$\rho(\mathbf{X}_n) = \ln \frac{\hat{P}(\mathbf{X}_n/R)}{\hat{P}(\mathbf{X}_n/\bar{R})} = \ln \frac{\hat{P}(R/\mathbf{X}_n)}{\hat{P}(\bar{R}/\mathbf{X}_n)} - \ln \frac{\hat{P}(R)}{\hat{P}(\bar{R})} . \tag{3.5.26}$$

Namely, we measure the effect of past experience $\mathbf{X}_n$ on the progress of learn-

ing by the increment of the quantity $\ln(\hat{P}(R/\mathbf{X}_n)/\hat{P}(\bar{R}/\mathbf{X}_n))$ from the value at the state of no experience, i.e., at $t = 0$.

The recurrence relation (3.5.18) can be easily grasped by the use of quantification ρ, i.e., it is equivalent to the following equality.

$$\rho(\mathbf{X}_n) = \rho(\mathbf{X}_{n-1}) + \rho(X_n), \tag{3.5.27}$$

where

$$\begin{aligned}\rho(X_n) &= \ln \frac{\hat{P}(X_n/R)}{\hat{P}(X_n/\bar{R})} \\ &= \ln \frac{P_n(F_n)\cdot\hat{P}(G_n/F_n, R)}{P_n(F_n)\cdot\hat{P}(G_n/F_n, \bar{R})} \\ &= \ln \frac{\hat{P}(G_n/F_n, R)}{\hat{P}(G_n/F_n, \bar{R})} .\end{aligned} \tag{3.5.28}$$

Hence, the following additivity of ρ holds.

$$\rho(\mathbf{X}_n) = \sum_{t=1}^{n} \rho(X_t) \qquad \text{for } n \geq 1. \tag{3.5.29}$$

It should be noted that (3.5.29) does not mean the stochastic independence of $\rho(X_1)$, $\rho(X_2)$, ... , and $\rho(X_n)$. Because the random variable X_t is dependent upon X_1, X_2, ... , and X_{t-1}.

The quantification $\rho(\mathbf{X}_n)$ evaluates the degree of fitness of the hypothesis (sample type) R to the actual stochastic property of the random environment compared to $\bar{R}$. To pursue the behavior of the random variable $\rho(\mathbf{X}_n)$, we take note of the measure, the quantified mean $I(R|\bar{R} ; \tilde{P}(X_i))$, introduced in Section 3.2. The effect of action $F_n(=f_i)$ on $\rho(\mathbf{X}_n)$ is evaluated by

$$\begin{aligned}&I(R|\bar{R} ; P(X_n/F_n = f_i)) \\ &= E_{X_n}[\rho(X_n)/F_n=f_i] \\ &\quad E_{(F_n,G_n)}\left[\ln \frac{\hat{P}(G_n/F_n, R)}{\hat{P}(G_n/F_n, \bar{R})} / F_n = f_i \right] \\ &\quad P(G_n = g_1 / F_n = f_i) \ln \frac{\hat{P}(g_1/f_i, R)}{\hat{P}(g_1/f_i, \bar{R})} \\ &\qquad + P(G_n = g_2/F_n = f_i) \ln \frac{\hat{P}(g_2/f_i, R)}{\hat{P}(g_2/f_i, \bar{R})} \quad \text{for } i \quad 1,2.\end{aligned} \tag{3.5.30}$$

From (3.5.26), I is rewritten as

$$I(R|\bar{R}\ ;\ P(X_n/F_n = f_i)) = E_{X_n}[\rho(\mathbf{X}_n) \quad \rho(\mathbf{X}_{n-1})/F_n = f_i].$$

for i 1, 2. (3.5.31)

From (3.5.5) and (3.5.5'), the relation between action probability $P_n(f_i)$ and the measure I is

$$I(R|\bar{R}\ ;\ P(X_n/F_n \quad f_i))$$

$$E_{X_n}\left[\ \ln \frac{P_{n+1}(f_1)}{P_{n+1}(f_2)} \quad \ln \frac{P_n(f_1)}{P_n(f_2)} / F_n = f_i\right] \quad \text{for } i = 1, 2. \quad (3.5.32)$$

Classification of learning behavior by the quantified mean I

In the following, we consider the case where the random environment is stationary, i.e., $P(G_n = g_j/F_n = f_i)$ does not depend on n, and so is denoted by $P(g_j/f_i)$ for i = 1, 2 and j 1,2. Also, we denote $I(R|\bar{R}\ ;\ P(X_n/F_n = f_i))$ by $I(R|\bar{R}\ ;\ f_i)$ for i 1, 2.

Regarding $\{\rho(\mathbf{X}_t)\}$ as a random process, the signs of the above quantities have a connection with the martingale theory. In fact, Norman(1970) demonstrated that the behavior of the learning process of the beta model (which is equivalent to our model) can be classified into four types according to the signs of $I(R|\bar{R}\ ;\ f_1)$ and $I(R|\bar{R}\ ;\ f_2)$.

THEOREM(Norman) : Let $a_i = \text{Prob.}(\lim_{n\to\infty} P_n(f_i) = 1)$ for i = 1, 2, i.e., $a_1 = \text{Prob.}(\lim_{n\to\infty} \rho(\mathbf{X}_n) \quad \infty)$ and $a_2 = \text{Prob.}(\lim_{n\to\infty} \rho(\mathbf{X}_n) = -\infty)$. Then the following holds.

(a) If $I(R|\bar{R}\ ;\ f_1) > 0$ and $I(R|\bar{R}\ ;\ f_2) > 0$, then $a_1 = 1(a_2 \quad 0)$.

(b) If $I(R|\bar{R}\ ;\ f_1) < 0$ and $I(R|\bar{R}\ ;\ f_2) < 0$, then $a_2 = 1(a_1 = 0)$.

(c) If $I(R|\bar{R}\ ;\ f_1) > 0$ and $I(R|\bar{R}\ ;\ f_2) < 0$, then $a_1 > 0$, $a_2 > 0$, and $a_1 + a_2 = 1$. In addition, $a_1 \to 1$ as $\hat{P}(R) \to 1$ and $a_2 \to 1$ as $\hat{P}(R) \to 0$.

(d) If $I(R|\bar{R}\ ;\ f_1) < 0$ and $I(R|\bar{R}\ ;\ f_2) > 0$, then $a_1 \quad a_2 \quad 0$. Moreover, $\limsup_{n\to\infty} P_n(f_1) = 1$ and $\liminf_{n\to\infty} P_n(f_1) = 0$ (with probability 1).

Which of Norman's classification will occur depends on the determination of the subjective probabilities $\hat{P}(g_j/f_i, R)$ and $\hat{P}(g_j/f_i, \bar{R})$ (i 1,2 and j 1, 2) in connection with the objective probabilities $P(g_j/f_i)$ (i =1,2 and j 1,2), as discussed in the following.

Let ρ_{ij} denote $\rho(f_i, g_j) = \ln(\hat{P}(g_j/f_i, R)/\hat{P}(g_j/f_i, \bar{R}))$. Then, from (3.5.

30), we have

$$I(R|\bar{R} ; f_i) = (\rho_{i1} \quad \rho_{i2})P(g_1/f_i) + \rho_{i2}. \tag{3.5.30'}$$

Let k_i denote as follows.

$$k_i \quad \frac{- \rho_{i2}}{\rho_{i1} \quad \rho_{i2}} \tag{3.5.33}$$

It is obvious that

$$\rho_{i1} \cdot \rho_{i2} < 0 \tag{3.5.34}$$

It follows that

$$0 \leq k_i \leq 1, \tag{3.5.35}$$

$$\text{sgn } (\rho_{i1} \quad \rho_{i2}) = \text{sgn } (\rho_{i1}) \qquad \text{for } i = 1, 2. \tag{3.5.36}$$

From (3.5.8) and (3.5.9), we obtain

$$\text{if } \rho_{11} \geq 0, \text{ then } \rho_{21} > 0 \tag{3.5.37}$$

Therefore, the following three cases are possible.

$$\text{(i)} \quad \rho_{11} < 0, \ \rho_{21} < 0. \tag{3.5.38}$$

$$\text{(ii)} \quad \rho_{11} > 0, \ \rho_{21} > 0. \tag{3.5.39}$$

$$\text{(iii)} \quad \rho_{11} \leq 0, \ \rho_{21} \geq 0. \tag{3.5.40}$$

The above conditions are restated in terms of the subjective probabilities. Let $J(R)$ and $J(\bar{R})$ be intervals in $[0,1]$ as follows.

$$J(R) \triangleq [\hat{P}(g_1/f_1, R), \hat{P}(g_1/f_2, R)] \text{ , and} \tag{3.5.41}$$

$$J(\bar{R}) \triangleq [\hat{P}(g_1/f_2, \bar{R}), \hat{P}(g_1/f_1, \bar{R})], \tag{3.5.42}$$

where each ordering follows from (3.5.8) and (3.5.9) presented earlier. Then, the above three cases are

(i) $J(R) \cap J(\bar{R}) \quad \Phi$ and $j(\bar{R})$ is located to the right of $J(R)$

(cf. Fig. 3.5.2(i)) (3.5.38')

(ii) $J(R) \cap J(\bar{R}) = \Phi$ and $J(\bar{R})$ is located to the left of $J(R)$

(cf. Fig. 3.5.2(ii)) (3.5.39')

(iii) $J(R) \cap J(\bar{R}) \neq \Phi$ (cf. Fig. 3.5.2(iii)) (3.5.40')

From (3.5.30'), (3.5.33), (3.5.35) and (3.5.36), the Norman's classification for cases (i), (ii), and (iii) with respect to the objective probabilities (

penalty frequencies) $P(g_1/f_1)$ and $P(g_1/f_2)$ are as fhown in Fig. 3.5.3. That is, the learning behavior of our model with respect to the stochastic properties of the random environment is classified according to the threshold values k_1 and k_2. For example, in case (i), type (a) : $P_n(f_1) \to 1$ occures when the penalty frequencies caused by f_1 and f_2 are less than k_1 and k_2, respectively, and type (b) : $P_n(f_2) \to 1$ occurs when they are more than k_1 and k_2, respectively. Thus, in case (i), the differentiation of the learning process into (a) or (b) does not depend on the superiority of action f_1 to f_2 or that of f_2 to f_1, i.e., comparison between the values of $P(g_1/f_1)$ and $P(g_1/f_2)$. The original meaning of R and $\bar{R}$ represented by (3.5.8) and (3.5.9) suggests the behavior described by Fig. 3.5.4. Among the behaviors of cases (i), (ii) and (iii), case (iii) is the best approximation to the suggested behavior. Thus, in addition to the conditions on the subjective probabilities (3.5.8) and (3.5.9), we can adopt (3.5.40') as a reasonable condition. Condition (3.5.40') is rewritten as follows.

$$P(g_1/f_1, R) < P(g_1/f_1, \bar{R})$$

(hypothesis R looks with more favor on f_1 than $\bar{R}$ does),

$$P(g_1/f_2, R) > P(g_1/f_2, \bar{R})$$

(hypothesis $\bar{R}$ looks with more favor on f_2 than R does).

(3.5.40'')

In the following, let us consider some experimental results discussed by Bush, Galanter, and Luce(1959) to test the validity of beta model. They examined the results by Solomon and Wynne(1953), in which dogs were trained to jump a barrier to avoid an electric shock, and also the T-maze experiments by Galanter and Bush(1959), where rats were rewarded whenever they turned right, and were never rewarded when they turned left. Bush et al. calculated the parameter β_{ij} in these experiments. In the former experiment, let f_1 and f_2 be the avoidance action (jumping the barrier) and non-avoidance action (staying), respectively, and g_1 and g_2 be the penalty and non-penalty, respectively. Then the reported values are $\beta_{12} = 1.2$ and $\beta_{21} = 1.7$. In the experiment, the reinforcement rule was set such that f_1 was always followed by g_2 and f_2 was always followed by g_1; no data can be given to estimate the values of β_{11} and β_{22}. We have shown that case (iii) in (3.5.40') is in most accordance with common sense. As a typical situation of case (iii), we assume that $J(R)$ $J(\bar{R})$, which in turn results in $P(g_1/f_1, R) = P(g_1/f_2, \bar{R})$ and $P(g_1/f_2, R) = P(g_1/f_1, \bar{R})$. The above assumption says that the goodness of action f_1 (under

hypothesis A) and that of action f_2 (under hypothesis $\bar{R}$) are the same and the badness of f_1 is the same as that of f_2; the learning subjects have no information to presume superiority of one action to the other, beforehand (before the learning starts). The assumption is equivalent to the condition : β_{11} β_{12}^{-1} and $\beta_{22} = \beta_{21}^{-1}$. The assumption yields $P(g_1/f_1, R) = P(g_1/f_2, \bar{R})$ 0.19 and $P(g_1/f_2, R) = P(g_1/f_1, \bar{R})$ 0.33. Namely, the subjective probabilities of the penalty are considerably low under the hypotheses R and $\bar{R}$, and the actions f_1 and f_2. In this case, the critical values k_1 and k_2 coincide with each other and are calculated as 0.13. Thus we can say that the learning proceeds unidirectionally (case(a) or case(b) occurs) only if one of the objective probabilities $P(g_1/f_1)$ and $P(g_1/f_2)$ is less than 0.13. In other words, in order to proceed the learning unidirectionally, one of the actions, i.e. the avoidance action, should be followed by non-penalty (non-shock) with a frequency of more than 0.87.

In the latter experiment, let us denote g_1, g_2, f_1, and f_2 as non-reward, reward, turning to the right, and turning to the left, respectively. Also, in this case, the deterministic reinforcement rule was adopted such that f_1 was followed by g_2 and f_2 was followed by g_1. The experimental result was β_{12} 1.02 ~ 1.14 and β_{21} = 1.44 ~ 2.6. As a rough approximation, let us adopt the medians of the values as the estimation of β_{12} and β_{21}. Thus we obtain β_{12} = 1.08 and β_{21} 2.02. By the use of the same assumption as before, the values of the subjective probabilities are calculated as $P(g_1/f_1, R) = P(g_1/f_2, \bar{R})$ = 0.066 and $P(g_1/f_2, R) = P(g_1/f_1, \bar{R})$ = 0.135. That is to say, the subjective probabilities of reward are quite high irrespectively of the hypotheses and the actions. The critical values are calculated as $k_1 = k_2$ = 0.096. Therefore, in order to proceed the learning unidirectionally, the reward probability $P(g_2/f_1)$ should be set as being larger than 0.904.

Using the above data, Bush, Galanter, and Luce concluded that the effect of penalty is larger than that of non-penalty, and that of non-reward is larger than that of reward. These conclusions are interpreted, through our model, as the subjective probabilities of penalty and those of non-reward being less than 0.5. Namely, if it is the case, then the likelihood ratio $P(g_1/f_2, R)/P(g_1/f_2, \bar{R})$ of the penalty or non-reward is larger than the likelihood ratio $P(g_2/f_1, R)/P(g_2/f_1, \bar{R})$ of non-penalty or reward. Since, we have $P(g_1/f_2, R)/P(g_1/f_2, \bar{R})$ $1 + [(P(g_1/f_2, R)$ $P(g_1/f_2, \bar{R}))/P(g_1/f_2, \bar{R})] > 1 + [(P(g_1/f_2, R)$ $P(g_1/f_2, \bar{R}))/(1$ $P(g_1/f_2, R))] = P(g_2/f_1, R)/P(g_2/f_1, \bar{R})$. In the above, the inequality and the last equality follow from the above mentioned fact and the assumption, respectively.

Interpretation of learning behavior as Markov chain

As shown in (3.5.19), response probability $P_{n+1}(f_i)$ is completely determined by its previous value $P_n(f_i)$ and action-reinforcement pair X_n at t n, for i = 1, 2. Hence, the quantity $\rho(\mathbf{X}_n) = \ln(\hat{P}(\mathbf{X}_n/R)/\hat{P}(\mathbf{X}_n)/\bar{R}))$ ln $(P_{n+1}(f_1)/P_{n+1}(f_2)) + \ln(\hat{P}(R)/\hat{P}(\bar{R}))$ represents all the necessary information contained in past experience $\mathbf{X}_n$ with respect to the future behavior of the learning process. Regarding the above quantity $\rho(\mathbf{X}_n)$ as a *state variable* at t = n of the learning subject, the learning process constitutes a Markov chain with infinitely many states and Norman's classification can easily be grasped by examining the properties of the Markov chain. For simplicity, we confine ourselves to the case of symmetric subjective probabilities, i.e.,

$$\hat{P}(g_1/f_1, R) = \hat{P}(g_1/f_2, \bar{R}) = q,$$
$$\hat{P}(g_1/f_2, R) = \hat{P}(G_1/f_1, \bar{R}) = 1 \quad q,$$
$$0 < q < \frac{1}{2}. \tag{3.5.43}$$

In this case, conditions (3.5.8), (3.5.9) and (3.5.40'') on the subjective probabilities are satisfied. From (3.5.28), (3.5.29), (3.5.31) and (3.5.43), the possible numerical values of $\rho(\mathbf{X}_n)$ (n = 0, 1, 2, ···) are given by

$$s_k \quad k \cdot \ln \frac{1 - q}{q} + \rho(\mathbf{X}_0) \qquad \text{for } k = 0, \pm 1, \pm 2, \ldots \tag{3.5.44}$$

where,

$$\rho(\mathbf{X}_n) \quad \ln \frac{\hat{P}(R)}{\hat{P}(\bar{R})} \tag{3.5.45}$$

The setting of the initial state variable $\rho(\mathbf{X}_0)$ has no important effect on the asymptotic behavior of $\{\rho(\mathbf{X}_n)\}$ under Norman's classification. Hence, without loss of generality, we make the assumption that

$$\hat{P}(R) = \hat{P}(\bar{R}) = \frac{1}{2}, \text{ i.e., } \rho(\mathbf{X}_0) \quad s_0 = 0. \tag{3.5.46}$$

Thus the states are distributed on the real axis with the same interval ln ((1 q)/q) as shown in Fig. 3.5.5. From (3.5.31) and (3.5.27), the transition from a state s_k is only possible to its adjacent states s_{k+1} or s_{k-1};

$$s_k \rightarrow s_{k+1} \text{ if } F_n \quad f_1 \text{ and } G_n = g_2$$
$$\text{or } F_n = f_2 \text{ and } G_n = g_1$$

and

$$s_k \rightarrow s_{k-1} \text{ if } F_n = f_1 \text{ and } G_n = g_1$$

$$\text{or } F_n = f_2 \text{ and } G_n = g_2,$$

$$\text{for any } k \text{ and } n\ (\geq 1). \tag{3.5.47}$$

From the definition of state: $\rho(\mathbf{X}_n) = \ln\left(P_{n+1}(f_1)/P_{n+1}(f_2)\right)$ $\ln\left(P_{n+1}(f_1)/(1\right.$ $\left.P_{n+1}(f_1)\right)$, the response probability at each state s_k is

$$P(f_1/s_k) = \exp(s_k)/(1 + \exp(s_k)),$$

$$P(f_2/s_k) = 1 \quad P(f_1/s_k) = 1/(1 + \exp(s_k)),$$

$$\text{for any } k. \tag{3.5.48}$$

Consequently, as shown in Fig. 3.5.6, the process $\{\rho(\mathbf{X}_n)\}$ constitutes a Markov chain with transition probabilities $P_{s_k s_{k+1}}$ and $P_{s_k s_{k-1}}$:

$$P_{s_k s_{k+1}} \quad P(f_1/s_k)\cdot P(g_2/f_1) + P(f_2/s_k)\cdot P(g_1/f_2),$$

$$P_{s_k s_{k-1}} \quad P(f_1/s_k)\cdot P(g_1/f_1) + P(f_2/s_k)\cdot P(g_2/f_2),$$

$$\text{for any } k. \tag{3.5.49}$$

Now, let us calculate the *average trend of the transitions* at each state. From (3.5.32), we have

$$I(R/\bar{R} : S_k) \triangleq E[\rho(\mathbf{X}_n) \quad \rho(\mathbf{X}_{n-1})/\rho(\mathbf{X}_{n-1}) = s_k]$$

$$= P(f_1/s_k)\cdot I(R|\bar{R};\ f_1) + P(f_2/s_k)\cdot I(R|\bar{R}\ ;\ f_2) \tag{3.5.50}$$

Since the quantity $P(f_1/s_k)(P(f_2/s_k))$ is monotonically increasing (decreasing) with respect to s_k, Norman's classification with respect to the above trend (3.5.50) becomes as follows.

$$\begin{aligned}
&\text{(a);} \quad I(R/\bar{R} : s_k) > 0 \quad \text{for arbitrary } s_k \\
&\text{(b);} \quad I(R/\bar{R} : s_k) < 0 \quad \text{for arbitrary } s_k \\
&\text{(c);} \quad I(R/\bar{R} : s_k) > 0 \quad \text{if } s_k > s_c^* \\
&\qquad\quad I(R/\bar{R} : s_k) < 0 \quad \text{if } s_k < s_c^* \\
&\text{(d);} \quad I(R/\bar{R} : s_k) < 0 \quad \text{if } s_k > s_d^* \\
&\qquad\quad I(R/\bar{R} : s_k) > 0 \quad \text{if } s_k < s_d^*
\end{aligned} \tag{3.5.51}$$

where the threshold values s_c^* and s_d^* are

$$\begin{aligned}
s_c^* &= \ln I(R|\bar{R}\ ;\ f_1) \quad (-\ln I(R|\bar{R}\ ;\ f_2)) \\
&= \ln I(R|\bar{R}\ ;\ f_1) \quad \ln I(\bar{R}|R\ ;\ f_2), \\
s_d^* &= \ln\left(-I(R|\bar{R}\ ;\ f_1)\right) \quad \ln I(R|\bar{R}\ ;\ f_2)
\end{aligned} \tag{3.5.52}$$

$$= \ln I(\bar{R}|R\ ;\ f_1) \quad \ln I(R|\bar{R}\ ;\ f_2).$$

Fig. 3.5.7 shows the above classification, where each arrow represents the direction and the amount of trend $I(R/\bar{R} : s_k)$ of the transitions at each state s_k. By the figure, Norman's results that $\rho(\mathbf{X}_n) \to \infty$ w.p.1 (with probability 1) in case (a), $\rho(\mathbf{X}_n) \to -\infty$ w.p.1 in case (b), $\{\rho(\mathbf{X}_n)\}$ branches into two types: $\rho(\mathbf{X}_n) \to \infty$ and $\rho(\mathbf{X}_n) \to -\infty$ in case (c), $\{\rho(\mathbf{X}_n)\}$ oscillates infinitely many times from $-\infty$ to $+\infty$ in case (d), can be intuitively understood.

The behavior $\rho(\mathbf{X}_n) \to \infty$ in case (a) and $\rho(\mathbf{X}_n) \to -\infty$ in case (b) are in accordance with the claim (as indicated in Fig. 3.5.4) that the learning subject ultimately prefers the action which produces less frequency of penalty g_1 than the other.

Tsetlin(1973) proposed an interesting learning machine (automaton model) with linear chain structure similar to that described by (3.5.47). Incorporating the chain with left and right endpoints s_{-N} and s_N, he suppressed the occurrence of the undesirable behavior in cases (c) and (d). Thus, in his chain, the number of states is finite and the transition rule is the same as (3.5.47) except at the endpoints,i.e.,

$$\left.\begin{array}{l} s_k \to s_{k+1} \quad (\text{for } -N \le k < N) \\ s_N \to s_N \quad (\text{for } k = N) \end{array}\right\}$$

$$\text{if } F_n = f_1 \text{ and } G_n = g_2 \text{ or } F_n = f_2 \text{ and } G_n = g_1$$

and

$$\left.\begin{array}{l} s_k \to s_{k-1} \quad (\text{for } -N < k \le N) \\ s_{-N} \to s_{-N} \quad (\text{for } k = -N) \end{array}\right\}$$

$$\left.\text{if } F_n = f_1 \text{ and } G_n = g_1 \text{ or } F_n = f_2 \text{ and } G_n = g_2\right\} \quad (3.5.47')$$

Instead of our stochastic decision rule (3.5.48) of action, he adopted the following deterministic rule, which is derived by dichotomizing the value of (3.5.48) into 0 and 1 by the threshold value of 1/2, i.e.,

$$P(f_1/s_k) = \begin{cases} 1 & \text{if } s_k > 0, \text{ i.e., } 1 \le k \le N, \\ 0 & \text{if } s_k < 0, \text{ i.e., } -N \le k \le -1. \end{cases} \quad (3.5.48')$$

(In his chain the neutral state s_0 is omitted)

His learning machine has a very simple structure as shown in Fig. 3.5.8 and is called a *linear tactic* or $L_{2N,2}$. The role of the endpoints s_{-N} and s_N is to avoid the dissipation of the process to infinity (∞ or $-\infty$) in an undesirable

direction. Tsetlin evaluated the behavior of $L_{2N,2}$ by $\bar{W}(L_{2N,2})$, the average frequency of penalty g_1 caused by its actions and showed that

$$\lim_{N\to\infty} \bar{W}(L_{2N,2}) = \min \{P(g_1/f_1),\ P(g_1/f_2)\}. \tag{3.5.53}$$

The above means that $L_{2n,2}$ asymptotically attains the ideal behavior indicated by Fig. 3.5.4. Tsetlin's model, however, is not devised as a model of learning processes but merely as a learning machine to approximate the ideal behavior. Some generalizations of his model were made by Varshavskii & Vorontsova(1963), Hellman & Cover(1970 & 1971), and Baxa & Nolte(1972).

In the above, we examined the behavior of the learning process by means of the state variable $\rho(\mathbf{X}_n) = \ln (P_{n+1}(f_1)/(1 \quad P_{n+1}(f_1)))$. Next we consider the behavior of the response probability $P_n(f_1)$ $(= \hat{P}(R/\mathbf{X}_{n-1}))$ itself.

Martingale conditions of learning behavior

In this part, we consider the conditions under which the process $\{P_n\ (= P_n(f_1))\}$ constitutes a submartingale or a supermartingale together with the sequence of σ-fields $\{\sigma_n\ (= \sigma(\mathbf{X}_{n-1}))\}$. A random process $\{P_n\}$ is called a sub(super)martingale with respect to $\{\sigma_n\}$ provided that

(I) $\{P_n\}$ is adapted to $\{\sigma_n\}$, i.e., P_n is σ_n-measurable for $n \geq 1$.

(II) $E[|P_n|] < \infty$ for $n \geq 1$.

(III) $E[P_{n+1}/\sigma_n] \geq (\leq)\ P_n$ with probability 1, for $n \geq 1$

In our case, P_{n+1} is completely determined by P_n and Z_n. Hence, condition (III) is equivalent to

(III) $E[P_{n+1}/P_n] \geq (\leq)\ P_n$ with probability 1, for $n \geq 1$.

For details, refer to Doob(1953).

In general, it is very difficult to derive the condition analytically. So we confine our discussions to the case of symmetric subjective and objective probabilities, i.e., we make the assumption (3.5.43) and also the following:

$$P(g_1/f_1) = r,\ P(g_1/f_2) = 1 \quad r. \tag{3.5.54}$$

In this case, we have

$$k_1 = k_2 \quad \frac{1}{2}. \tag{3.5.55}$$

Therefore, the asymptotic behavior of $P_n(f_1)$ is

$$r < \frac{1}{2} \rightarrow P_n(f_1) \rightarrow 1 \quad \text{with probability 1,} \tag{3.5.56}$$

$$r > \frac{1}{2} \rightarrow P_n(f_1) \rightarrow 0 \quad (P_n(f_2) \rightarrow 1) \text{ with probability 1.} \tag{3.5.57}$$

From (3.5.5) and (3.5.5'), it is obvious that $\{P_n\}$ is adapted to $\{\sigma_n\}$. From the recurrence relations (3.5.5), (3.5.5'), (3.5.19) and (3.5.23), we obtain

$$E[P_{n+1}/P_n] = P_n[r\cdot\frac{\beta P_n}{\beta P_n + (1 - P_n)} + (1 \quad r)\cdot\frac{\beta^{-1}P_n}{\beta^{-1}P_n + (1 - P_n)}]$$

$$+ (1 \quad P_n)[(1 \quad r)\cdot\frac{\beta^{-1}P_n}{\beta^{-1}P_n + (1 \quad P_n)} + r\cdot\frac{\beta P_n}{\beta P_n + (1 \quad P_n)}], \tag{3.5.58}$$

where

$$\beta = \frac{q}{1 \quad q} < 1. \tag{3.5.59}$$

This leads to

$$E[P_{n+1}/P_n] \quad P_n = \frac{(1 - P_n)((2 \quad \beta \quad \beta^{-1})P_n - \{1 \quad (\beta r + \beta^{-1} \quad r\beta^{-1})\})}{(\beta P_n + (1 \quad P_n))(\beta P_n + (1 - P_n))}. \tag{3.5.60}$$

It is obvious for arbitrary values of P_n and $\beta(<1)$ that $1 \quad P_n > 0$, $2 \quad \beta \quad \beta^{-1} < 0$, $\beta P_n + (1 \quad P_n) > 0$, $\beta^{-1}P_n + (1 \quad P_n) > 0$ and $\beta^{-1} \quad \beta > 0$. Thus, we obtain that $E[P_{n+1}/P_n] \geq P_n$ for $0 \leq P_n \leq 1$ if and only if the following holds.

$$1 \quad (\beta r + \beta^{-1} \quad r\beta^{-1}) \leq 2 \quad \beta \quad \beta^{-1}. \tag{3.5.61}$$

Similarly, we obtain that $E[P_{n+1}/P_n] \leq P_n$ for $0 \leq P_n \leq 1$ if and only if the following holds.

$$1 \quad (\beta r + \beta^{-1} \quad r\beta^{-1}) \geq 0. \tag{3.5.62}$$

The above two inequalities (3.5.61) and (3.5.62) are reduced as follows.

$$r \leq \frac{\beta}{1 + \beta} = \frac{\frac{q}{1 \quad q}}{1 + \frac{q}{1 \quad q}} = q. \tag{3.5.61'}$$

$$r \geq \frac{1}{1 + \beta} = 1 \quad q. \tag{3.5.62'}$$

Threrfore, the following theorem holds.

THEOREM 3.5.1 : In the symmetric case conditioned by (3.5.43) and (3.5.54), the process $\{P_n(f_1)\}$ is a submartingale if and only if $r \leq q$, and also $\{P_n(f_1)\}$ is a supermartingale if and only if $r \geq 1 \quad q$.

Comparing (3.5.56) with (3.5.61'), and also (3.5.57) with (3.5.62'), these martingale conditions are stronger than Norman's conditions (cf. Fig. 3.5.9). The above theorem says that the necessary and sufficient condition that

$\{P_n(f_1)\}$ has a martingale property is

$$\left.\begin{aligned} J(R) = J(\bar{R}) &\subset [P(g_1/f_1),\ P(g_1/f_2)],\ \text{if } r < \frac{1}{2} \\ J(R) \quad J(\bar{R}) &\subset [P(g_1/f_2),\ P(g_1/f_1)],\ \text{if } r > \frac{1}{2} \end{aligned}\right\}. \tag{3.5.63}$$

Fig. 3.5.10. shows the above conditions, i.e., in order that the learning process has the martingale property, the difference between the subjective probabilities for f_1 and f_2 conditioned upon R or $\bar{R}$ should not be greater than that of the objective (actual) probabilities. In other words, the subjective probabilities should be set moderately compared with the actual situation. From (3.5.61') and (3.5.62'), the value of ρ_{ij} (= ln $P(g_j/f_i, R)$ ln $P(g_j/f_i, \bar{R})$, i 1, 2 and j = 1, 2), which represents the difference between hypothesese R and $\bar{R}$, has an upper bound ln ((1 r)/r) (if $r < \frac{1}{2}$) or ln (r/(1 r)) (if $r > \frac{1}{2}$) and a lower bound ln (r/(1 r)) (if $r < \frac{1}{2}$) or ln ((1 r)/r) (if $r > \frac{1}{2}$). Consequently, the values of the quantified mean $I(R|\bar{R}\ ;\ f_i)$ (i 1, 2), the expected values of the above quantities, has an upper bound (1 2r) ln ((1 r)/r) and a lower bound (1 2r) ln (r/(1 r)). Therefore, from (3.5.31) and (3.5.32), the learning processes with the martingale property behave conservatively and their learning speed is limited.

When P_n (= $P_n(f_1)$) constitutes a sub(super)martingale, as indicated by the property (III) of the martingale condition, the frequency of action f_1 is, on the average, monotonically increasing (decreasing), i.e., the learning process has an immutable trend. Among the monotone properties of sub(super)martingales, we refer to Dubins' inequality, which is a refined version of the famous Doob's inequality and seems to be worth noting to examine the behavior of learning processes. The inequality concerns the frequency of oscillations in positive supermartingale processes. If $\{P_n(f_1)\}$ is a submartingale, then $\{1 \quad P_n(f_1)\}$ constitutes a positive supermartingale. Therefore, the Dubins' inequality can be restated as follows (For details refer to Dubins(1962)).

THEOREM(Dubins) : Let $\bar{\gamma}_{ab}$ ($\underline{\gamma}_{ab}$) be the number or up crossings (down crossings) of the interval [a,b] by the process $\{P_n(f_1)\}$, i.e., the number of times the process $\{P_n(f_1)\}$ passes from below a (above b) to above b(below a,), for $0 < a < b < 1$. Then the following statements hold.

(1) If $\{P_n(f_1)\}$ is a supermartingale, then

$$P(\bar{\gamma}_{ab} \geq k) \leq (\frac{a}{b})^k \cdot \min\{\frac{\hat{P}(R)}{a},\ 1\}$$

(2) If $\{P_n(f_1)\}$ is a submartingale, then

$$P(\gamma_{ab} \geq k) \leq \left(\frac{1 - b}{1 \quad a}\right)^k \cdot \min\left\{\frac{\hat{P}(\bar{R})}{1 \quad b}, 1\right\}$$

Hence, the behaviors of the learning processes are not so often opposite to the immutable direction.

3.5.3 Evaluation of Memory Functions on the Learning Behavior based on the Quantification Method

In the previous sections, we only treated the case of learning subject without memory functions, i.e., the hypotheses R and $\bar{R}$ are described by conditional probabilities $\hat{P}(g_j/f_i, R)$ and $\hat{P}(g_j/f_i, \bar{R})$ for i = 1, 2 and j 1, 2. In this section, we consider the case of a learning subject with memory functions. In this case, as aforementioned at the end of condition 2 on the subjective probabilities in Section 3.5.2, the hypotheses R and $\bar{R}$ (at t = n) are described by conditional probabilities $\hat{P}(g_j/f_i, R, \mathbf{X}'_{n-1})$ and $\hat{P}(g_j/f_i, \bar{R}, \mathbf{X}'_{n-1})$, where $\mathbf{X}'_{n-1}$ is a subsequence of past experience $\mathbf{X}_{n-1}$, i.e., the memorized experience by the learning subject. Let π denote the selection function of $\mathbf{X}'_{n-1}$ from $\mathbf{X}_{n-1}$, i.e.,

$$\mathbf{X}'_{n-1} = \pi(\mathbf{X}_{n-1}). \tag{3.5.64}$$

The learning subject only memorizes $\pi(\mathbf{X}_{n-1})$ out of $\mathbf{X}_{n-1}$, and sets up subjective probabilities under R and $\bar{R}$ as $\hat{P}(g_j/f_i, R, \pi(\mathbf{X}_{n-1}))$ and $\hat{P}(g_j/f_i, \bar{R}, \pi(\mathbf{X}_{n-1}))$. That is to say, (3.5.10) and (3.5.11) in condition 2) of section 3.5.1 are altered as follows:

$$\hat{P}(G_n = g_j/F_n = f_i, R, \mathbf{X}_{n-1}) \quad \hat{P}(g_j/f_i, R, \pi(\mathbf{X}_{n-1})), \tag{3.5.10'}$$

$$\hat{P}(G_n = g_j/F_n = f_i, \bar{R}, \mathbf{X}_{n-1}) = \hat{P}(g_j/f_i, \bar{R}, \pi(\mathbf{X}_{n-1})). \tag{3.5.11'}$$

Comparing with (3.5.10) and (3.5.11) (no memory case), the learning subject has the ability to alter the subjective probabilities according to the selected past experience $\mathbf{X}'_{n-1}$. In this case, the revision of the *a posteriori* probabilities of the hypotheses are as follows. From (3.5.10'), (3.5.17) is altered to (3.5.17'):

$$\hat{P}(R, G_n = g_j/F_n = f_i, \mathbf{X}_{n-1}) = \hat{P}(g_j/f_i, R, \pi(\mathbf{X}_{n-1})) \cdot \hat{P}(R/\mathbf{X}_{n-1}). \tag{3.5.17'}$$

Consequently, recurrence relation (3.5.18) is altered to (3.5.18'):

$$\hat{P}(R/\mathbf{X}_n)$$

$$= \frac{\hat{P}(g_j/f_i,\ R,\ \pi(\mathbf{X}_{n-1}))\cdot\hat{P}(R/\mathbf{X}_{n-1})}{\hat{P}(g_j/f_i,\ R,\ \pi(\mathbf{X}_{n-1}))\cdot\hat{P}(R/\mathbf{X}_{n-1}) + \hat{P}(g_j/f_i,\ \bar{R},\ \pi(\mathbf{X}_{n-1}))\cdot\hat{P}(\bar{R}/\mathbf{X}_{n-1})},$$

$$\hat{P}(\bar{R}/\mathbf{X}_n) = 1 \quad \hat{P}(R/\mathbf{X}_n)$$

when $F_n = f_i$ and $G_n = g_j$ for $i = 1, 2$ and $j = 1, 2$. (3.5.18')

Therefore, the learning model in this case is also equivalent to the beta model in which coefficient β_{ij} is not constant but depends on $\pi(\mathbf{X}_{n-1})$ for i 1, 2 and j 1, 2, i.e.,

$$\beta_{ij}(\pi(\mathbf{X}_{n-1})) = \frac{\hat{P}(g_j/f_i,\ R,\ \pi(\mathbf{X}_{n-1}))}{\hat{P}(g_j/f_i,\ \bar{R},\ \pi(\mathbf{X}_{n-1}))}\ . \tag{3.5.25'}$$

In the following, we evaluate the effect of memory functions on the learning process by the use of the measure I.

For simplicity, we assume that actions f_1 and f_2 do not intervene with each other, i.e., f_1 does not affect the stochastic property of the random environment for action f_2 and also *vice versa*. Also, we assume that the stochastic properties of the random environment are described by stationary simple Markov chains; i.e., the output sequence of the enrironment whose input sequence is a repetition of f_1s (a repetition of f_2s) constitutes a stationary simple Markov chain. In this case, output G_n depends not only on input F_n but also on the output corresponding to the last input which is the same as F_n. Therefore, we naturally set the following simple selection function π:

$$\pi(\mathbf{X}_{n-1}) = (G_{f_1},\ G_{f_2})\ , \tag{3.5.65}$$

where G_{f_i} is the output corresponding to the last f_i input for i = 1, 2; for example, $G_{f_1} = g_2$ and G_{f_2} g_1 in the case of Fig. 3.5.11. From the supposition that f_1 and f_2 do not interfere with each other, the subjective probabilities in (3.5.11') and (3.5.12'), i.e., $\hat{P}(g_j/f_i,\ R,\ \pi(\mathbf{X}_{n-1}))$ and $\hat{P}(g_j/f_i,\ \bar{R},\ \pi(\mathbf{X}_{n-1}))$, can be replaced by $\hat{P}(g_j/f_i,\ R,\ G_{f_i})$ and $\hat{P}(g_j/f_i,\ \bar{R},\ G_{f_i})$, respectively, for i 1, 2.

In this section, we discuss the effect of the memory functions based on the quantity $I(R|\bar{R}\ ;\ f_i)$ introduced in Section 3.5.2.

Let us consider the case $F_n = f_1$. That is, we measure the efficiency of learning with memory against that of learning without memory by the use of

$I(R|\overline{R}\ ;\ f_1)$. Let r_{uv} be the transition probability of the Markov chain under action f_1, i.e.,

$$r_{uv} = P(G_n = g_v/G_{f_1} \quad g_u), \quad u = 1, 2 \text{ and } v = 1, 2. \tag{3.5.66}$$

When the chain is an absorbing Markov chain, the output sequence becomes a repetition of the same reinforcement g_1 (or g_2) after falling into the absorbing state. Threfore, we assume that the chain is regular, i.e.,

$$r_{uv} > 0 \quad \text{for } u = 1, 2 \text{ and } v \quad 1, 2. \tag{3.5.67}$$

The corresponding transition diagram is shown in Fig. 3.5.12.

As the learning subjects, we consider two types, a learning subject L_0 without memory and a learning subject L_1 with a memory, as follows. The hypotheses R_0 and $\overline{R}_0$ of the subject L_0 are set as similar to (3.5.43);

(a) L_0; hypotheses R_0 and $\overline{R}_0$ for action f_1

$$\begin{aligned} &R_0;\ \hat{P}(g_1/f_1,\ R_0) = q,\ \hat{P}(g_2/f_1,\ R_0) \quad 1 \quad q. \\ &\overline{R}_0;\ \hat{P}(g_1/f_1,\ \overline{R}_0) = 1 \quad q,\ \hat{P}(g_2/f_1,\ \overline{R}_0) = q. \end{aligned} \tag{3.5.68}$$

The subject L_1 with a memory has the hypotheses R_1 and $\overline{R}_1$ as follows.

(b) L_1; hypotheses R_1 and $\overline{R}_1$ for action f_1

$$R_1;\ \hat{P}(g_j/f_1,\ R_1,\ g_{j'}) = \begin{cases} q & \text{if } j \neq j', \\ 1 \quad q & \text{if } j = j', \end{cases}$$

$$\overline{R}_1;\ \hat{P}(g_j/f_1,\ \overline{R}_1,\ g_{j'}) = \begin{cases} 1 \quad q & \text{if } j \neq j', \\ q & \text{if } j \quad j', \end{cases}$$

$$\text{for } j = 1, 2 \text{ and } j' = 1, 2. \tag{3.5.69}$$

Compared with the learning subject L_0, the subject L_1 has the option to use the hypothesis R_0 or $\overline{R}_0$ according to whether the last output caused by f_1 was g_1 or g_2. That is, the hypothesis R_1 corresponds to using hypothesis $\overline{R}_0$ or R_0 according to the last output being g_1 or g_2, respectively. Also, $\overline{R}_1$ corresponds to using R_0 or $\overline{R}_0$according to the last output being g_1 or g_2, respectively. In this sence, L_1 can be regarded as having a memory function compared with L_0 which uses the hypotheses R_0 and $\overline{R}_0$ no matter what the last output is. In the following, we compare $I(R_0|\overline{R}_0\ ;\ f_1)$ with $I'(R_1|\overline{R}_1\ ;\ f_1)$ which denotes a natural extension of I for the learning subjects with memory functions.

In this case that $F_n = f_1$, $I(R_0|\overline{R}_0\ ;\ f_1)$ in (3.5.30) is represented as

follows.

$$I(R_0|\bar{R}_0 \ ; \ f_1) = E_{G_n}\left[\ln \frac{\hat{P}(G_n/f_1, R_0)}{\hat{P}(G_n/f_1, \bar{R}_0)}\right]$$

$$\left(= E_{G_{f_1}}\left[E_{G_n}\left[\ln \frac{\hat{P}(G_n/f_1, R_0)}{\hat{P}(G_n/f_1, R_0)} /G_{f_1}\right]\right] \right). \qquad (3.5.70)$$

For the efficiency of L_1, we introduce the following quantity $I'(R_1|\bar{R}_1 \ ; \ f_1)$ as the natural extension of the above measure.

$$I'(R_1|\bar{R}_1 \ ; \ f_1) = E_{G_{f_1}}\left[E_{G_n}\left[\ln \frac{\hat{P}(G_n/f_1, R_1, G_{f_1})}{\hat{P}(G_n/f_1, \bar{R}_1, G_{f_1})} /G_{f_1}\right]\right]. \qquad (3.5.71)$$

To calculate the measures indicated by (3.5.70) and (3.5.71), we need the values of $P(G_n)$ and $P(G_{f_1})$. From the supposition that the Markov chain under action f_1 is regular, $P(G_n)$ and $P(G_{f_1})$ tend to the stationary probability disribution along with the number of the trials by f_1 irrespective of the initial state. Therefore, we assume that $P(G_n)$ and $P(G_{f_1})$ are equal to the stationary probability distribution (p_1^*, p_2^*), i.e.,

$$P(G_n = g_j) \quad P(G_{f_1} = g_j) - p_j^* \quad \text{for } j = 1, 2. \qquad (3.5.72)$$

where p_1^* and p_2^* satisfy the following stationary distribution condition:

$$p_1^* \cdot r_{11} + p_2^* \cdot r_{21} = p_1^*, \quad (p_1^* + p_2^* = 1). \qquad (3.5.73)$$

Thus, we obtain

$$\left.\begin{aligned} P(G_n = g_1) = P(G_{f_1} = g_1) = \frac{r_{21}}{r_{12} + r_{21}} \\ P(G_n = g_2) = P(G_{f_1} = g_2) = \frac{r_{12}}{r_{12} + r_{21}} \end{aligned}\right\}. \qquad (3.5.74)$$

From (3.5.74) and (3.5.68), we have

$$I(R_0|\bar{R}_0 \ ; \ f_1) = \frac{r_{12} \quad r_{21}}{r_{12} + r_{21}} \ln \frac{1 \quad q}{q}. \qquad (3.5.75)$$

From (3.5.69), we have

$$E_{G_n}\left[\ \ln \frac{\hat{P}(G_n/f_1,\ R_1,\ G_{f_1})}{\hat{P}(G_n/f_1,\ \bar{R}_1,\ G_{f_1})}\ /G_{f_1}\right] = \begin{cases}(r_{11} - r_{12})\ln \dfrac{1-q}{q} & \text{if } G_{f_1} = g_1,\\ (r_{22} - r_{21})\ln \dfrac{1-q}{q} & \text{if } G_{f_1} = g_2.\end{cases} \tag{3.5.76}$$

Then, from (3.5.74), (3.5.71) is calculated as follows:

$$I'(R_1|\bar{R}_1\ ;\ f_1) = \frac{r_{12} + r_{21} - 4r_{12}r_{21}}{r_{12} + r_{21}} \ln \frac{1-q}{q}\ . \tag{3.5.77}$$

Let us compare the efficiency of models L_0 and L_1 by the use of the information theoretic measure I and I'. As shown in (3.5.75) and (3.5.77), I and I' contain hypothetical probability q in the same form $\ln((1-q)/q)$; that form changes sign according to whether $q < \frac{1}{2}$ or not. Hence, q has no essential meaning with respect to the comparison. In the following, we consider the case of $q < \frac{1}{2}$. This means, in case of model L_0, hypothesis R_0 presumes that the penalty frequency of the environment $P(g_1/f_1)$ ($= p_1^* = r_{12}/(r_{12} + r_{21})$) is less than 1/2 and $\bar{R}_0$ presumes that the frequency is larger than 1/2. On the other hand, in case of model L_1, hypothesis R_1 presumes that r_{12} and r_{21}, the probabilities of two succeeding outputs being unequal, are less than 1/2, and $\bar{R}_1$ presumes that the probabilities are larger than 1/2. The sign of I(or I') represents the average direction of reinforcements, i.e., which of R_0 or $\bar{R}_0$ (which of R_1 or $\bar{R}_1$) is reinforced by a trial of f_1. The absolute value of I (or I') evaluates the degree of reinforcement or average speed of learning. Fig. 3.5.13 shows the signs of I and I', and also the relation between the absolute values of I and I'. In the figure, I is positive in regions (5) (8) and negative in (1) (4), I' is positive in (1), (2), (7), and (8), and negative in (3) (6). The borderline of positive and negative regions for I is represented by $r_{12} = r_{21}$ and that for I' is represented by $r_{21} = r_{12}/(4r_{12} - 1)$. Hence, if p_1^* is less (larger) than 1/2, then R_0 ($\bar{R}_0$) is reinforced, and if both r_{12} and r_{21} are less (larger) than 1/2, then R_1 ($\bar{R}_1$) is reinforced. These statement are in accordance with the presumptions of hypotheses R_0, $\bar{R}_0$, R_1 and $\bar{R}_1$ mentioned before. In Fig. 3.5.13, we can see that $|I'| > |I|$ in regions (1), (8), (4) and (5). That is, if the stochastic properties of the environment are in accordance with the presumption of hypothesis R_1 such as in regions (1) and (8) (i.e., $0 < r_{12} < \frac{1}{2}$ and $0 < r_{21} < \frac{1}{2}$), or are in accordance with those of hypothesis $\bar{R}_1$ such as in regions (4) and (5) ($\frac{1}{2} < r_{12} < 1$ and $\frac{1}{2} < r_{21} < 1$), then the model with a memory function, L_1, has a higher

efficiency than the model without a memory function, L_0. However, in the other regions, L_0 is more efficient than L_1. Hence, the hypotheses R_1 and $\overline{R}_1$ should be set up in accordance with the stochastic properties of the environment in order to make the memory functions effective.

It should be noted that the above discussion was only on action f_1. However, we can develop a similar discussion based on f_2 by setting up the Markov chain property of the random environment and the hypotheses of L_0 and L_1, under f_2.

3.5.4 On Some Extensions of the Learning Model

In previous sections, we only treated the case in which the learning subject has two hypotheses R and $\overline{R}$. In this section, we extend our model in such a way that the learning subject has more than two hypotheses or a continuum of hypotheses. For simplicity, we confine the discussions to the case of a learning subjects without memory. However, discussions concerning the case with memory functions can be analyzed in a similar manner.

Case of a learning subject having finite multiple hypotheses

Let us assume that the learning subject has h hypotheses R_1, R_2, ... , and R_h. We denote by $\hat{P}(g_j/f_i, R_\nu)$ the subjective probability of output g_j conditioned upon input f_i and hypothesis R_ν for i 1, 2, j 1, 2, and ν 1, 2, ... , h. From the relation that $\hat{P}(g_2/f_i, R_\nu) = 1 \quad \hat{P}(g_1/f_i, R_\nu)$, each hypothesis R_ν is charactirized by the pair of quantities $\hat{P}(g_1/f_1, R_\nu)$ and $\hat{P}(g_1/f_2, R_\nu)$, i.e., by a point e_ν in the plane $E = [0,1] \times [0,1]$. Thus, all the hypotheses R_1, R_2, ... , and R_h are characterized by the h points e_1, e_2, ... , and e_h in E as shown in Fig. 3.5.14.

To give the decision rule of action, we assign each hypothesis R_ν to one of the actions f_1 and f_2 as similar to the case of two hypotheses R and $\overline{R}$. That is, we set two regions E_1 and E_2 in E such that

$$E_1 \cup E_2 = E = [0, 1] \times [0, 1] ,$$
$$E_1 \cap E_2 = \Phi . \tag{3.5.78}$$

If e_ν is in region E_i, then hypothesis R_ν is assigned to f_i for i = 1, 2 and ν 1, 2, ... , h. For example, in accordance with the ideal behavior in Fig. 3.5.4, the regions are set as shown in Fig. 3.5.14. The decision of the next action in this case is also given by a random experiment (coin tossing)

with the following probability law (instead of (3.5.5), (3.5.5'), (3.5.6) and (3.5.6')).

$$P_{n+1}(f_i) = P(F_{n+1} = f_i/\mathbf{X}_n) = \sum_{e_\nu \in E_i} \hat{P}(R_\nu/\mathbf{X}_n)$$

$$\text{for } n \geq 1 \text{ and } i = 1, 2, \qquad (3.5.5'')$$

$$P_1(f_i) = \sum_{e_\nu \in E_i} \hat{P}(R_\nu) \qquad \text{for } i = 1, 2. \qquad (3.5.6'')$$

where $\hat{P}(R_\nu/\mathbf{X}_n)$ is the *a posteriori* subjective probability of R_ν conditioned upon past experience $\mathbf{X}_n$ and $\hat{P}(R_\nu)$ is the *a priori* subjective probability of R_ν, for $\nu = 1, 2, \ldots, h$.

Renewal of the *a posteriori* subjective probabilities is given as similar to (3.5.18), i.e.,

$$\hat{P}(R_\nu/\mathbf{X}_n) = \frac{\hat{P}(g_j/f_i, R_\nu)\cdot\hat{P}(R_\nu/\mathbf{X}_{n-1})}{\sum_{\mu=1}^{h} \hat{P}(g_j/f_i, R_\mu)\cdot\hat{P}(R_\mu/\mathbf{X}_{n-1})}$$

$$\text{when } X_n = (f_i, g_j). \qquad (3.5.18'')$$

From (3.5.5'') and (3.5.18''), our model is equivalent to the beta model only when h 2, i.e., the case of two hypotheses, where $P_{n+1}(f_1)$ and $P_{n+1}(f_2)$ depend only on $P_n(f_1)$ and $P_n(f_2)$ and X_n as shown in Section 3.5.1. In general, Sternberg(1963) called the models with the following property *path independent* models.

$$P_{n+1}(f_i) = w_i(P_n(f_1), P_n(f_2), X_n) \qquad \text{for } i = 1, 2.$$

$$\text{(where } w_1 \text{ and } w_2 \text{ are arbitrary functions.)} \qquad (3.5.79)$$

If h > 2, $P_{n+1}(f_1)$ and $P_{n+1}(f_2)$ are determined not only by $P_n(f_1)$, $P_n(f_2)$ and X_n but also by the values of $\hat{P}(R_1/\mathbf{X}_{n-1})$, $\hat{P}(R_2/\mathbf{X}_{n-1})$, ... , and $\hat{P}(R_h/\mathbf{X}_{n-1})$. Hence, the case of two hypotheses without memory is path independent, while the case with more than two hypotheses and the case with memory function are path dependent.

In the case of multiple (more than two) hypotheses (multiple sample types), the quantification ρ is defined for each pair of sample types R_ν and $R_{\nu'}$ as follows:

$$\rho(R_\nu/R_{\nu'}\ ;\ \mathbf{X}_n) \triangleq \ln \frac{\hat{P}(\mathbf{X}_n/R_\nu)}{\hat{P}(\mathbf{X}_n/R_{\nu'})}$$

$$= \ln \frac{\hat{P}(R_\nu/\mathbf{X}_n)}{\hat{P}(R_{\nu'}/\mathbf{X}_n)} \quad \ln \frac{\hat{P}(R_\nu)}{\hat{P}(R_{\nu'})}$$

$$\text{for } \nu, \nu' \quad 1, 2, \ldots, h. \tag{3.5.80}$$

From (3.5.18'), the recurrence relation (3.5.27) also holds in this case, i.e., we have

$$\rho(R_\nu/R_{\nu'}; \mathbf{X}_n) \quad \rho(R_\nu/R_{\nu'}; \mathbf{X}_{n-1}) + \rho(R_\nu/R_{\nu'}; X_n) \tag{3.5.27'}$$

Hence, the additivity (3.5.29) of is also valid;

$$\rho(R_\nu/R_{\nu'}; \mathbf{X}_n) = \sum_{t=1}^{n} \rho(R_\nu/R_{\nu'}; X_t) . \tag{3.5.29'}$$

From the above and (3.5.80), the *a posteriori* probabilities are given as follows :

$$\hat{P}(R_\nu/\mathbf{X}_n) \quad \frac{\hat{P}(R_\nu) \cdot \prod_{t=1}^{n} \hat{P}(G_t/F_t, R_\nu)}{\sum_{\mu=1}^{h} \hat{P}(R_\mu) \cdot \prod_{t=1}^{n} \hat{P}(G_t/F_t, R_\mu)}$$

$$\text{for } \nu = 1, 2, \ldots, h. \tag{3.5.81}$$

Case of a learning subject having a continuum of hypotheses

The above discussions can be easily extended to the case of continuum of hypotheses. In this case, any hypothesis R is represented as a point (ξ,η) in $E = [0, 1] \times [0, 1]$, where $\xi \quad \hat{P}(g_1/f_1, R)$ and $\eta = \hat{P}(g_1/f_2, R)$ and R is denoted by $R_{\xi\eta}$ (cf. Fig. 3.5.15). So, $\hat{P}(R_{\xi\eta}/\mathbf{X}_n)$ denotes *a posteriori* subjective probability density of $R_{\xi\eta}$ conditioned upon $\mathbf{X}_n$ and satisfies the following recurrence formula as similar to (3,5,18').

$$\hat{P}(R_{\xi\eta}/\mathbf{X}_n) = \frac{\hat{P}(g_j/f_i, R_{\xi\eta}) \cdot \hat{P}(R_{\xi\eta}/\mathbf{X}_{n-1})}{\iint_E \hat{P}(g_j/f_i, R_{\xi\eta}) \cdot \hat{P}(R_{\xi\eta}/\mathbf{X}_{n-1}) d\xi d\eta}$$

$$\text{for } 0 < \xi < 1 \quad \text{and } 0 < \eta < 1. \tag{3.5.18'''}$$

Using a similar method, (3.5.81) is extended as the following.

$$\hat{P}(R_{\xi\eta}/\mathbf{X}_n) = \frac{\hat{P}(R_{\xi\eta}) \cdot \prod_{t=1}^{n} \hat{P}(G_t/F_t, R_{\xi\eta})}{\iint_E \hat{P}(R_{\xi\eta}) \cdot \prod_{t=1}^{n} \hat{P}(G_t/F_t, R_{\xi\eta}) d\xi d\eta}$$

$$\frac{\hat{P}(R_{\xi\eta})\xi^{n_{11}}(1 \quad \xi)^{n_{12}}\eta^{n_{21}}(1 \quad \eta)^{n_{22}}}{\iint_E \hat{P}(R_{\xi\eta})\xi^{n_{11}}(1 \quad \xi)^{n_{12}}\eta^{n_{21}}(1 \quad \eta)^{n_{22}}d\xi d\eta}, \qquad (3.5.81')$$

where n_{ij} is the number of occurrence of the pair (f_i, g_j) in $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ for i 1, 2 and j = 1, 2, and $\hat{P}(R_{\xi\eta})$ is the *a priori* subjective probability density of $R_{\xi\eta}$.

Laming(1969) also derived the above beta type distribution in his study on choice-reaction time experiments. Eq. (3.5.81') is an extension of his result into a two-parameter case.

Similarly, the response probability is given by

$$P_{n+1}(f_i) = \iint_{E_i} \hat{P}(R_{\xi\eta}/\mathbf{X}_n)d\xi d\eta \qquad \text{for } n \geq 1, \qquad (3.5.5''')$$

$$P_1(f_i) = \iint_{E_i} \hat{P}(R_{\xi\eta})d\xi d\eta , \qquad (3.5.6''')$$

where E_i is the subregion of E corresponding to action f_i for i = 1, 2.

As it is difficult to calculate the above integral, we confine ourselves to the symmetric case which is an extension of (3.5.43), i.e.,

$$\hat{P}(R_{\xi\eta}) = 0 \quad \text{if } \eta \neq 1 \quad \xi, \qquad \text{for } 0 \leq \xi \leq 1. \qquad (3.5.43')$$

more precisely,

$$\hat{P}(R_{\xi\eta}) = \delta(\eta \quad (1 \quad \xi)), \qquad (3.5.43'')$$

where δ is Dirac's delta function.

Let R_ξ denote $R_{\xi(1-\xi)}$. Then R_ξ has the following initial distribution.

$$\hat{P}(R_\xi) = 1 \qquad \text{for } 0 \leq \xi \leq 1. \qquad (3.5.43''')$$

In this case, (3.5.81') is

$$\hat{P}(R_\xi/\mathbf{X}_n) \quad \frac{\xi^{n_2}(1 \quad \xi)^{n_1}}{\int_0^1 \xi^{n_2}(1 - \xi)^{n_1}d\xi} \quad \frac{(n+1)!}{n_1!n_2!}\xi^{n_2}(1 \quad \xi)^{n_1}, \qquad (3.5.81'')$$

where

$$n_1 \quad n_{12} + n_{21}, \quad n_2 \quad n_{11} + n_{22}, \quad (n = n_1 + n_2), \qquad (3.5.82)$$

that is, n_1 represents the number of favorable trials for f_1; the sum of n_{12}(the number of non-penalized trials on which action f_1 occurs) and n_{21}(the

number of penalized trials on which f_2 occurs). The larger n_1 is, the more f_1 is supposed to be the favorable action. Similar statements hold for n_2 and f_2. From (3.5.81''), these numerals are sufficient statistics of $\mathbf{X}_n$. Hence, we denote $\hat{P}(R_\xi/\mathbf{X}_n)$ by $\hat{P}(R_\xi/n_1, n_2)$. If the regions E_1 and E_2 are set up as shown in Fig. 3.5.15, then (3.5.5''') is

$$P_{n+1}(f_1/n_1, n_2) = \frac{(n+1)!}{n_1!n_2!}\int_0^{\frac{1}{2}}\xi^{n_2}(1-\xi)^{n_1}d\xi \ . \tag{3.5.83}$$

Using the method of integration by parts, we have

$$P_{n+1}(f_1/n_1, n_2) = P_{n+1}(f_1/n_1+1, n_2+1) + (\tfrac{1}{2})^{n+1}\,{}_{n+1}C_{n_1} \ . \tag{3.5.84}$$

From simple calculations,

$$\left.\begin{aligned} P_{n+1}(f_1/0, n) &= (\tfrac{1}{2})^{n+1} \\ P_{n+1}(f_1/n, 0) &= 1 - (\tfrac{1}{2})^{n+1} \end{aligned}\right\}. \tag{3.5.85}$$

Consequently, we obtain

$$\left.\begin{aligned} P_{n+1}(f_1/n_1, n_2) &= (\tfrac{1}{2})^{n+1}({}_{n+1}C_0 + {}_{n+1}C_1 + \cdots + {}_{n+1}C_{n_1}) \\ &= \text{Prob.}(Y_{n+1} \le n) = \text{Prob.}(Y_{n+1} > n_2), \\ P_{n+1}(f_2/n_1, n_2) &= \text{Prob.}(Y_{n+1} \le n_2) = \text{Prob.}(Y_{n+1} > n_1), \end{aligned}\right\}. \tag{3.5.86}$$

where Y_{n+1} denotes the number of heads in n+1 succeeding coin tossings with equal probabilities of heads and tails. To interpret the meaning of (3.5.86), suppose that the learning subject faces the following fictitious random environment.

$$P(g_1/f_1) = P(g_1/f_2) = \frac{1}{2} \ . \tag{3.5.87}$$

That is to say, the environment takes a neutral attitude to f_1 and f_2. Then, Y_{n+1} can be interpreted as n_1', the number of favorable outcomes for f_1 in trials of n+1 times from the neutral environment. Decision rule (3.5.86) means that if the result n_1' is less than or equal to the actual result n_1, i.e., actual result n_1 is more favorable for action f_1 than (or equal to) result n_1' from the neutral environment, then action f_1 is used, and if the actual result is less favorable for f_1 than the result from neutral environment, then f_2 is used.

From the well known relation on binomial coefficients that

$${}_uC_v = {}_{u-1}C_v + {}_{u-1}C_{v-1} \qquad \text{for } 1 \le v < u \ , \tag{3.5.88}$$

the revision of response probability is given as follows.

$$\left.\begin{aligned} P_{n+2}(f_1/n_1 + 1,\ n_2) &= P_{n+1}(f_1/n_1,\ n_2) + (\tfrac{1}{2})^{n+2}\ {}_{n+1}C_{n_1+1} \\ P_{n+2}(f_1/n_1,\ n_2 + 1) &= P_{n+1}(f_1/n_1,\ n_2) \quad (\tfrac{1}{2})^{n+2}\ {}_{n+1}C_{n_1} \end{aligned}\right\}. \tag{3.5.89}$$

From (3.5.84), we finally obtain

$$P_{n+1}(f_1/n_1,\ n_2) = \tfrac{1}{2}[P_n(f_1/n_1 \quad 1,\ n_2) + P_n(f_1/n_1,\ n_2 - 1)]$$

$$\text{for } n_1 \geq 1 \text{ and } n_2 \geq 1. \tag{3.5.90}$$

The above averaging property (3.5.90) and (3.5.85) describe the behavior of the learning subject having a continuum of hypotheses under symmetric subjective probabilities indicated by (3.5.43''). From the relation that $P_{n+1}(f_1/n_1,\ n_2) + P_{n+1}(f_2/n_1,\ n_2) = 1$, (3.5.90) is equivalent to the following.

$$P_{n+1}(f_1/n_1,\ n_2) \quad P_n(f_1/n_1 \quad 1,\ n_2)$$

$$= P_{n+1}(f_2/n_1,\ n_2) \quad P_n(f_2/n_1,\ n_2 \quad 1)\ . \tag{3.5.91}$$

The above relation means that the model has a kind of symmetry on its reinforcements, i.e., the reinforcement of f_1 by a trial $(f_1,\ g_2)$ or $(f_2,\ g_1)$ is the same as that of f_2 by a trial $(f_2,\ g_2)$ or $(f_1,\ g_1)$. Fig. 3.5.16 shows the values of $P_{n+1}(f_1/n_1,\ n_2)$ for $n \leq 5$. In the figure, (3.5.91) means that a b = b c. It is obvious from the figure that the incremental value $\Delta P_n = P_{n+1}(f_1/n_1,\ n_2) \quad P_n(f_1/n_1 \quad 1,\ n_2)$ has its maximum value at $P_{n+1} = 1/2$ or $P_n \quad 1/2$, for each n. From (3.5.89), we have

$$\Delta P_n \quad (\tfrac{1}{2})^{n+1}\ {}_nC_{n_1}\ . \tag{3.5.92}$$

Let us assume that n is even number. Then P_n has its maximum value at $n_1 = n/2$, i.e.,

$$(\Delta P_n)_{max} = (\tfrac{1}{2})^{n+1}\ {}_nC_{\frac{n}{2}} = (\tfrac{1}{2})^{n+1}\frac{n!}{((\frac{n}{2})!)^2}\ . \tag{3.5.93}$$

From Stirling formula that

$$n! \sim \sqrt{2\pi}\ e^{-n} n^{n+\frac{1}{2}}, \tag{3.5.94}$$

we obtain

$$(\Delta P_n)_{max} \sim (\tfrac{1}{2})^{n+1}\sqrt{2\pi}e^{-n}n^{n+\frac{1}{2}} \cdot (\sqrt{2\pi}e^{-\frac{n}{2}} \cdot (\tfrac{n}{2})^{\frac{n}{2}+\frac{1}{2}})^{-2} = (2\pi)^{-\frac{1}{2}} \cdot n^{-\frac{1}{2}}. \tag{3.5.95}$$

Hence, the maximum value of ΔP_n has an order of $O(n^{-\frac{1}{2}})$.

On the other hand, in the case of two hypotheses R and $\bar{R}$ with symmetric subjective probabilities indicated by (3.5.43), ΔP_n has its maximum value $(1 - \sqrt{\beta})/(1 + \sqrt{\beta})$ at $P_n = \sqrt{\beta}/(1 + \sqrt{\beta}) < 1/2$ where β is given by (3.5.59). Also, (3.5.90) does not hold in this case. In fact, let $P \triangleq P_{n+1}(f_1/n_1, n_2)$, $P' \triangleq P_n(f_1/n_1 \quad 1, n_2)$ and $P'' \triangleq P_n(f_1/n_1, n_2 - 1)$. Then, from (3.5.19), we have

$$\frac{P}{1 \quad P} = \beta^{-1} \cdot \frac{P'}{1 \quad P'} = \beta \cdot \frac{P''}{1 \quad P''} , \tag{3.5.96}$$

it follows that

$$\frac{P}{1 \quad P} = \sqrt{\frac{P'}{1 \quad P'} \frac{P''}{1 \quad P''}} . \tag{3.5.97}$$

Hence, P has a nonlinear relation with P' and P'' as follows (cf. (3.5.90)):

$$P = \frac{\sqrt{P'P''(1 - P')(1 - P'')} \quad P'P''}{1 \quad P' \quad P''} . \tag{3.5.98}$$

Moreover, let $t' \quad P' \quad \frac{1}{2}$ and $t'' = P'' \quad \frac{1}{2}$. Then, we have

$$P = -\frac{1}{4} \frac{\sqrt{(1 - 4t'^2)(1 - 4t''^2)} - (2t' + 1)(2t'' + 1)}{t' + t''} . \tag{3.5.99}$$

From Taylor expansion,

$$\sqrt{1 \quad 4u^2} = 1 \quad 2u^2 \quad 2u^4 + O(u^6) \qquad \text{for } 0 \le u < 1/2, \tag{3.5.100}$$

it follows that

$$\begin{aligned} P = {} & \frac{1}{2}(t' + t'' + 1) + \frac{1}{2}(t' + t'')(t' \quad t'') + O(t'^3 \cdot t''^2) \\ & + O(t'^2 \cdot t''^3) \\ & \frac{1}{2}(P' + P'') + \frac{1}{2}(P' + P'' \quad 1)(P' \quad P'')^2 \\ & + O((P' \quad \tfrac{1}{2})^3 \cdot (P'' \quad \tfrac{1}{2})^2) + O((P' \quad \tfrac{1}{2})^2 (P'' \quad \tfrac{1}{2})^3) . \end{aligned} \tag{3.5.101}$$

Thus, equality (3.5.90) holds approximately only when $P' \cong P'' \cong 1/2$, i.e., only when $P_n(f_1/n_1 \quad 1, n_2) \cong 1/2$ and $\beta \cong 1$ (case of small learning).

In this section, it has been shown that the quantification ρ and the quantified mean I are closely related to the behavior of a kind of conditional learning process which is derived from randomized decision rules (3.5.5), (3.5.5'), (3.5.6), and (3.5.6') of the next action. It has also been shown that the learning process model is equivalent to the so-called Luce's beta model

under suppositions (3.5.10) and (3.5.11), while the two models are derived from completely different considerations. Relation (3.5.25) throws light on the concrete interpretation of the coefficient β in Luce's model. Based on the quantification ρ, it was shown that the asymptotic behavior of the model with respect to the objective probabilities of the environments has three cases (i), (ii) and (iii) as shown in Fig. 3.5.3, referring to Norman's theorem. As a result, condition (3.5.40'') on the subjective conditional probabilities is shown to be necessary together with conditions (3.5.8) and (3.5.9). Also, Markov chain properties of the learning behavior were examined in connection with Tsetlin's model. The essential difference of Tsetlin's model from ours is the existence of upper and lower bounds on the *a posteriori* subjective probabilities. The analysis by the martingale theory yielded that there is a limit on the speed of learning when the learning process has the desirable property called martingale property. By introducing a natural extension (3.5.71) of the quantified mean I, we compared the efficiency of a model having a memory function with that of a model having no memory functions. Eventually, it is difficult to expect that a model with a memory function has a higher efficiency than a model without memory functions under all stochastic properties of an environment. This is because a learning subject has only two hypotheses and, as a result, cannot cover all the possible stochastic properties of environments. In Section 3.5.4, it was shown that the models with more than two hypotheses or a continuum of hypotheses have quite different learning behaviors from those of the model with just two hypothese. The only property which is common throughout the models is the additivity of ρ expressed by (3.5.28) and (3.5.28').

3.6 Conclusions

In this chapter, we have considered aggregation problems of items and categories in response matrices (tables) in which each entity represents response probability and the number of the sample types are two, through quantifications for items and categories. In Section 3.2, we have introduced a quantification for each category in each item in response matrices by introducing reasonable crigeria (3.2.3), (3.2.4), and (3.2.5). Based on the above quantification, we introduced an information theoretic measure, called mean information intensity, of the degree of the separation between the two sample types. The fundamental properties of the above measure were stated by theorems 3.2.1 and 3.2.2. Also, in Section 3.3, its relationship with the

Bayesian discrimination rates in particular cases specified by (3.3.68) or (3.3.70) was shown by theorem 3.3.2. Some experiments by computer simulations assured the above relationship. However, in general situations, the Bayesian discrimination rates are not prescribed by the above measure but instead by the mean, called quantified mean, and the variance, called quantified variance of the above quantification as shown by (3.3.39). In Section 3.4, we have considered some efficient ways for aggregation of items and categories in response matrices by referring to the above measure, mean information intensity. In the item aggregation problems, the correlation measure between items defined by (3.4.2) plays an important role. For the category aggregation problems, we introduced a measure of information loss due to the aggregation defined by (3.4.109), and the problems were discussed in a more general framework such as the lattice structure of information. However, there is no such general framework for the item aggregation problems. Furthermore, it has been shown, in Section 3.5, that the quantification and the quantified mean are applicable to the analysis of the learning behavior of Luce's beta model, by introducing a conditional probability learning model whose behavior is equivalent to Luce's model.

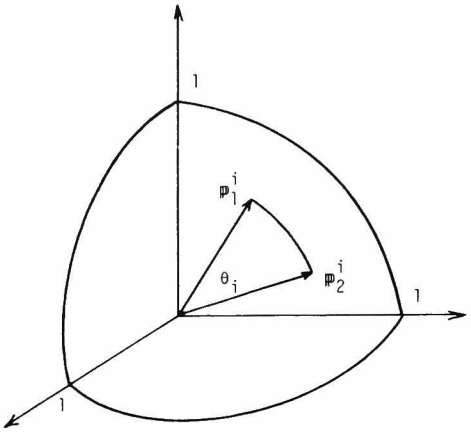


Fig. 3.3.1. Geometrical interpretation of Bhattacharyya distance.

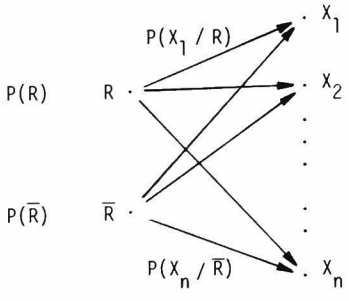


Fig. 3.3.2. Representation of items X_1, X_2, ···, and X_n as communication channels.

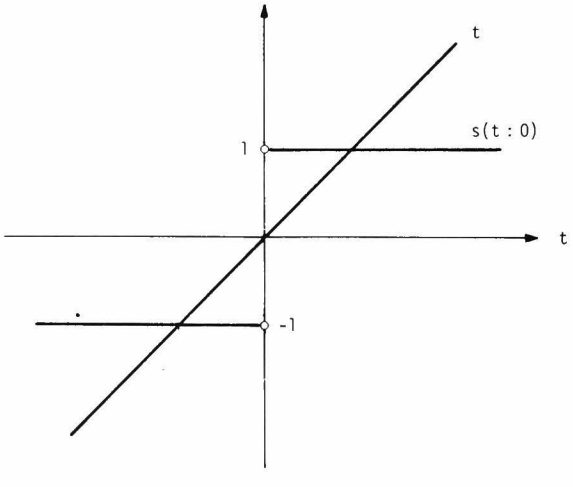


Fig. 3.3.3. Approximation of threshold function s(t : 0) by t.

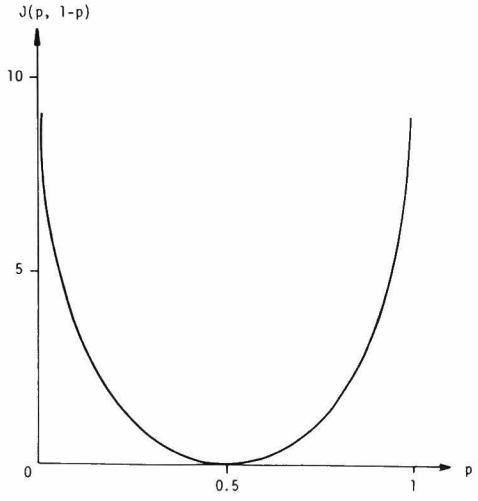


Fig. 3.3.4. The value of Hâjek's divergence J versus probability p.

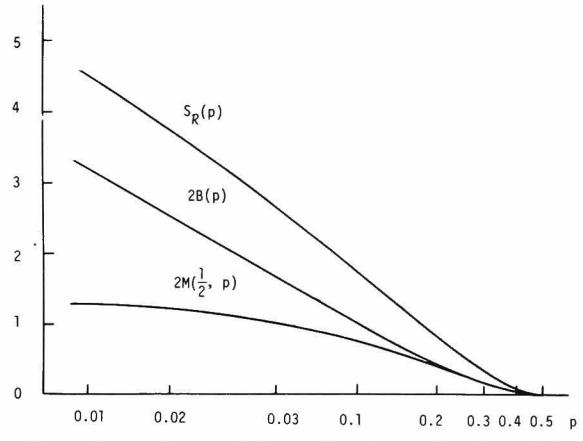


Fig. 3.3.5. The values of mean information intensity S_R and its components Bhattacharyya distance B and Shannon's mutual information M.

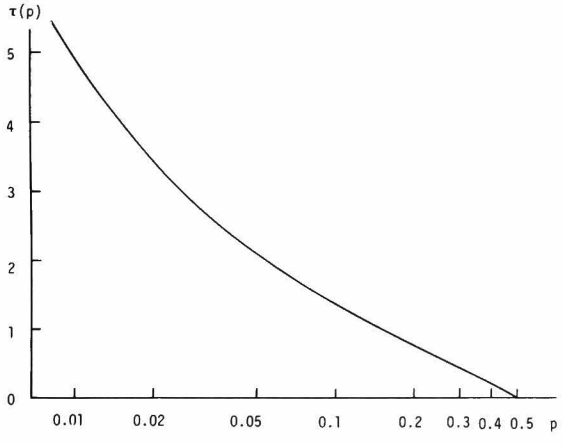


Fig. 3.3.6. The value of signal to noise ratio $\tau(p)$ of the random variable $\rho(X_i)$.

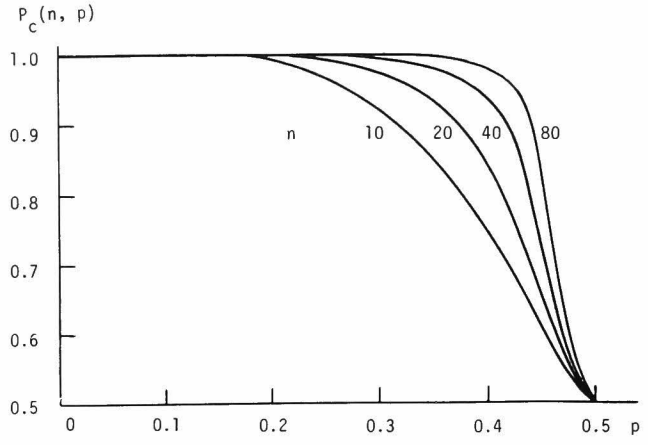


Fig. 3.3.7. The relation of discrimination rate $P_c(n, p)$ to the number n of items and the probability p.

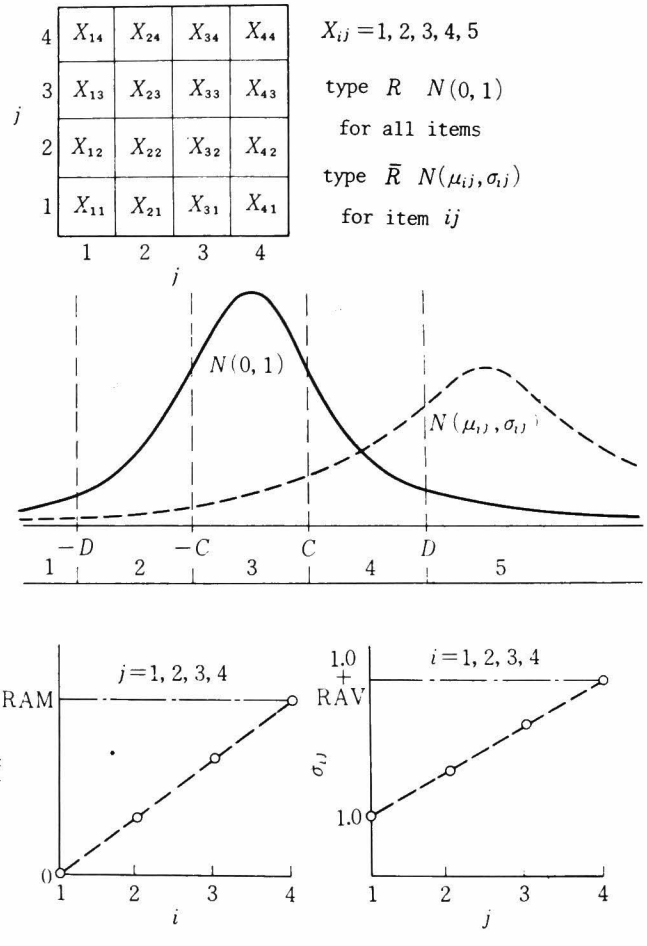


Fig. 3.3.8. Example model for computer simulation.

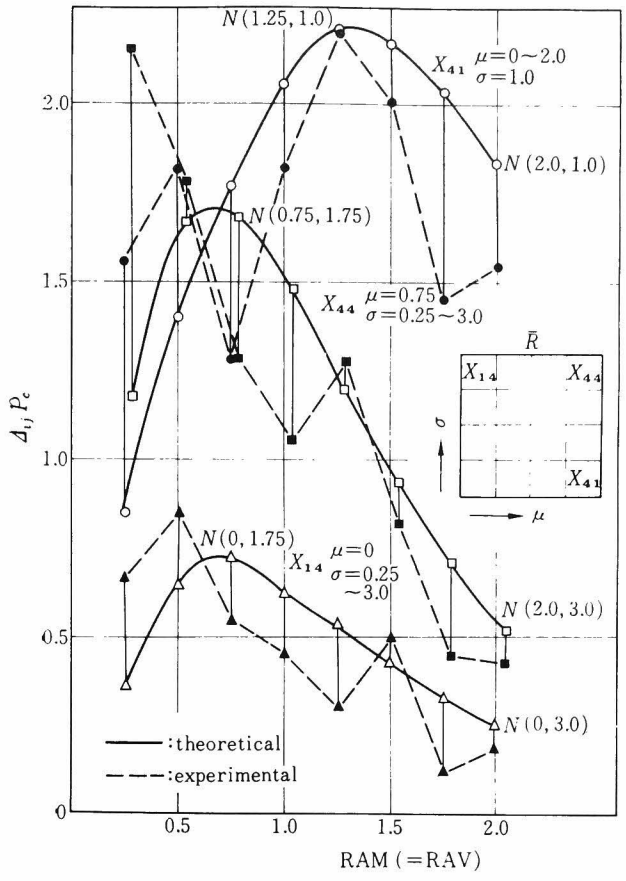


Fig. 3.3.9. Comparison of theoretical and experimental decrease of discrimination rates.

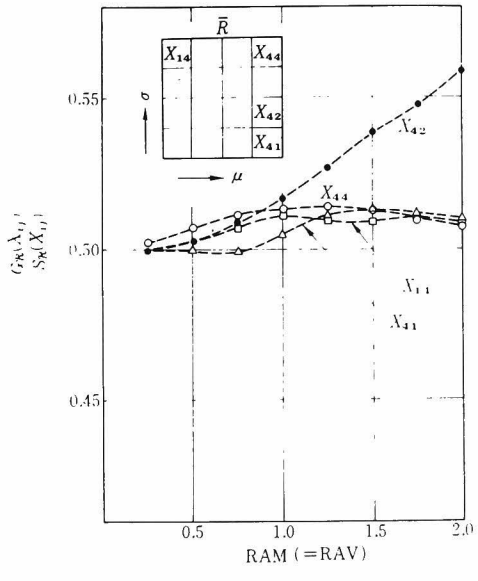


Fig. 3.3.10. Relation between the relative effectiveness G_R and the mean information intensity S_R.

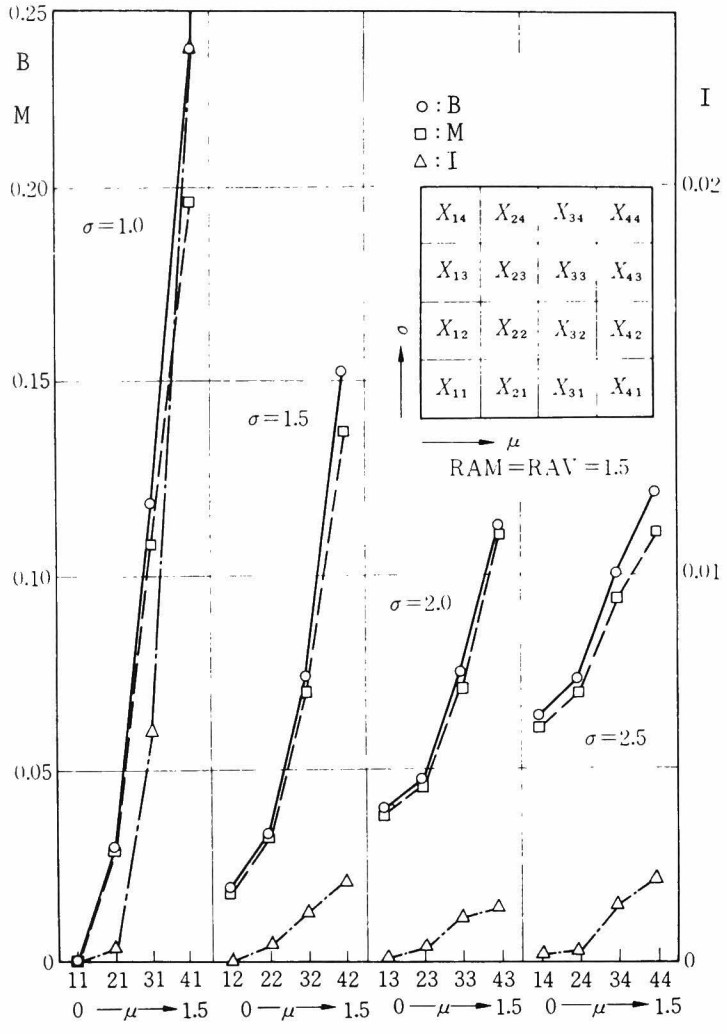


Fig. 3.3.11. Dependence of Bhattacharyya distance B, mutual information M and Kullback-Leibler information number I on the parameter μ and σ.

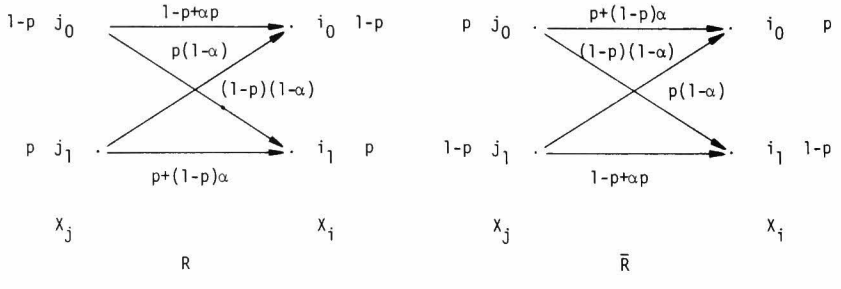


Fig. 3.4.1. Schematic representation of correlation between items X_i and X_j, where the degree of correlation is determined by the parameter α.

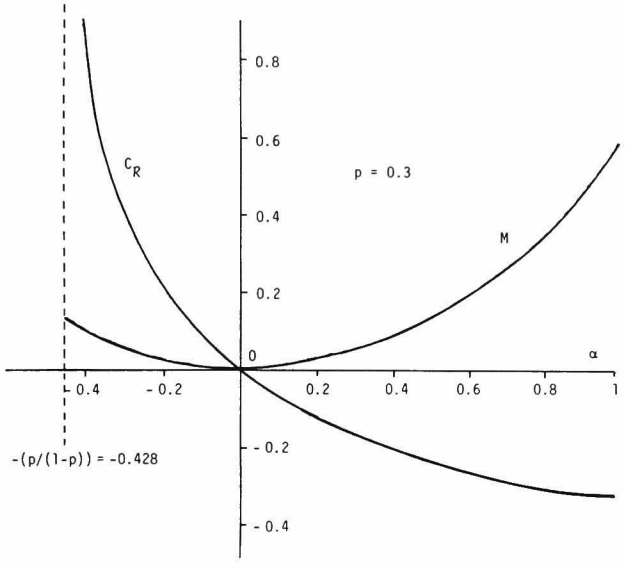


Fig. 3.4.2. The values of the measure C_R of the correlation and Shannon's mutual information M versus parameter α.

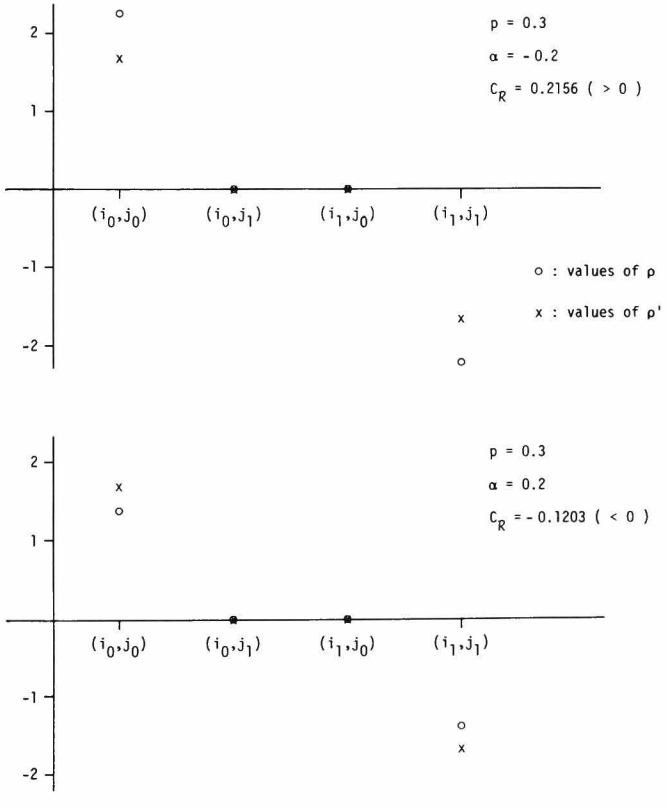


Fig. 3.4.3. The relation between ρ and ρ' under $C_R > 0$ and $C_R < 0$.

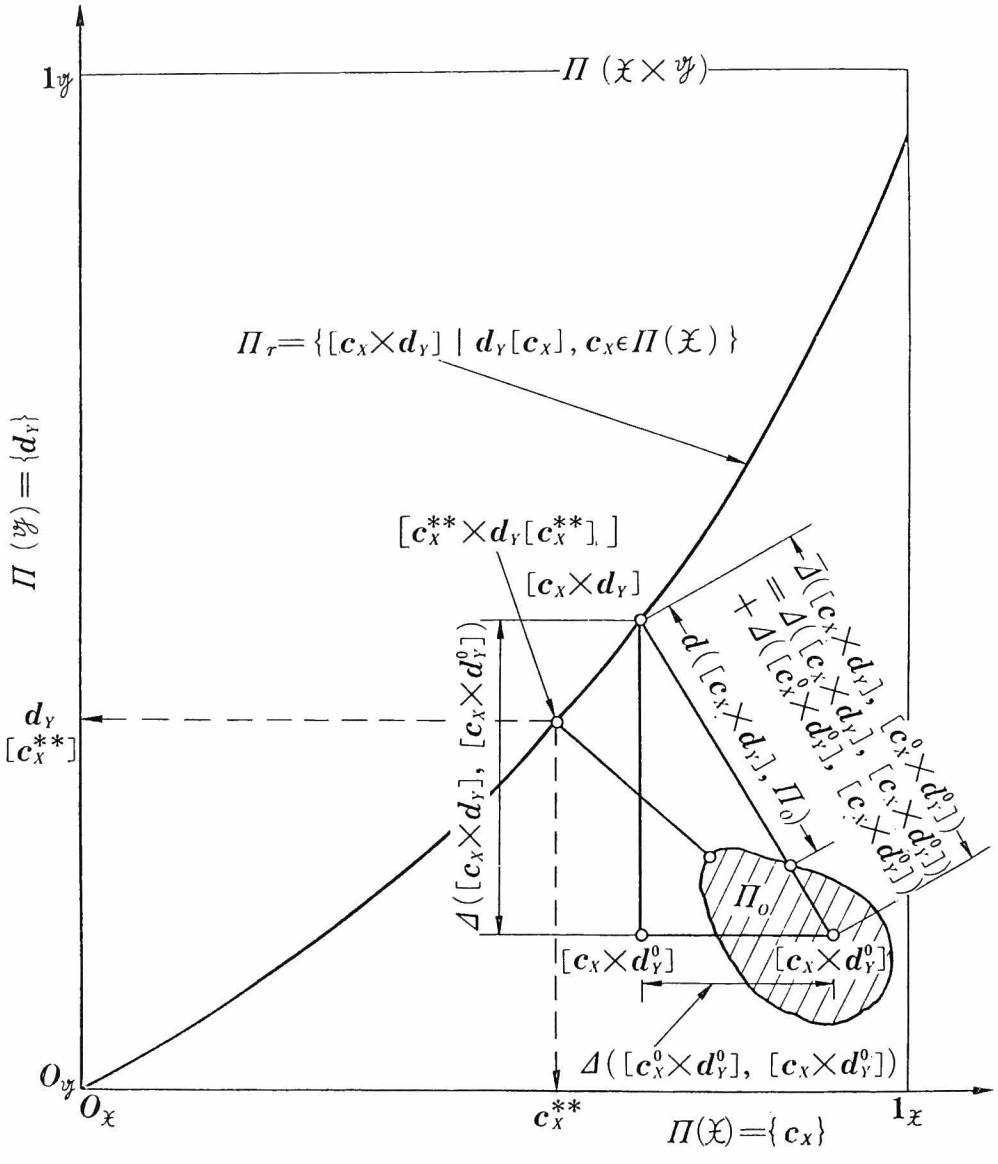


Fig. 3.4.4. Determination of suboptimum compression c_X^{**} of input information.

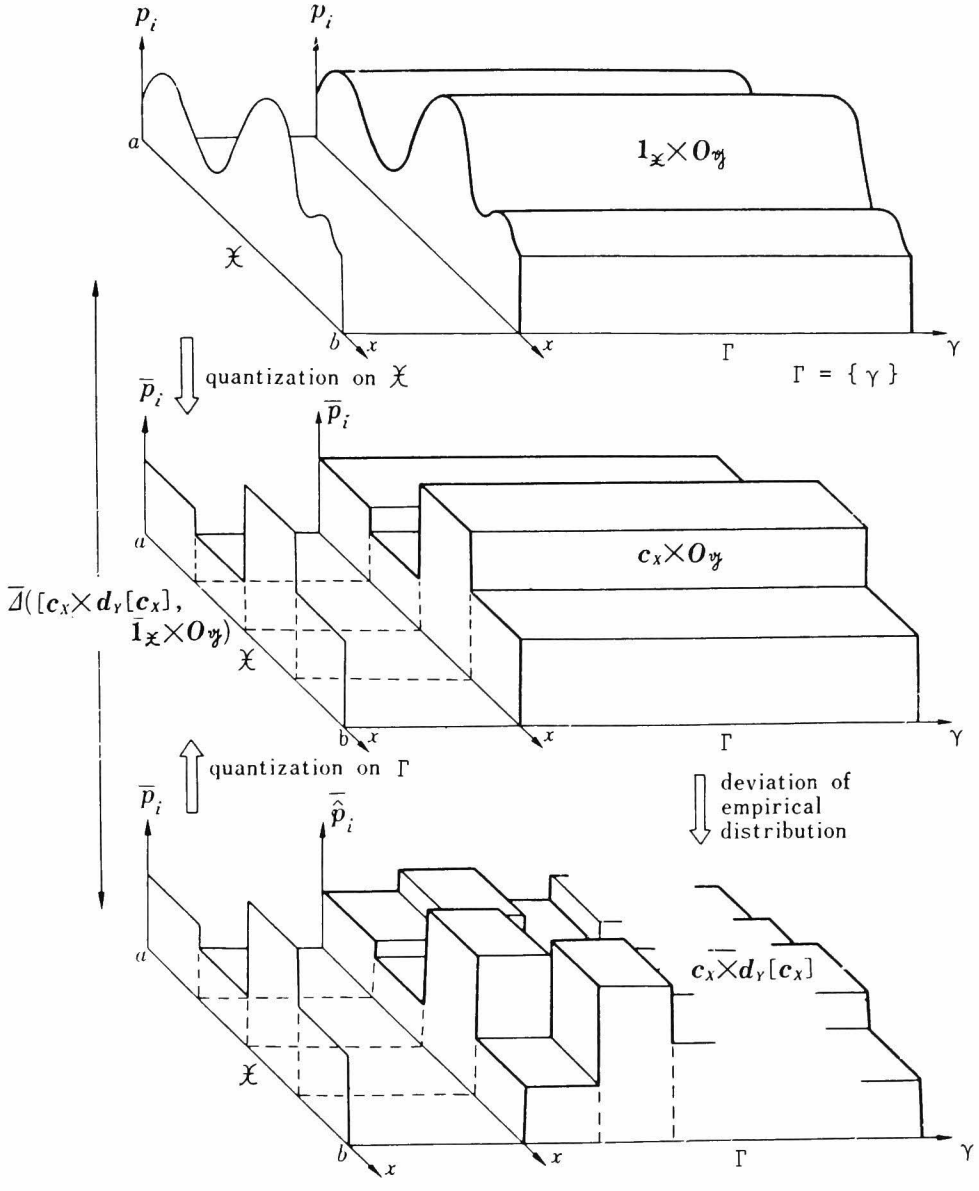


Fig. 3.4.5. Relation among original, quantized and empirical distributions in the quantization problem of continuum of categories.

HOGY

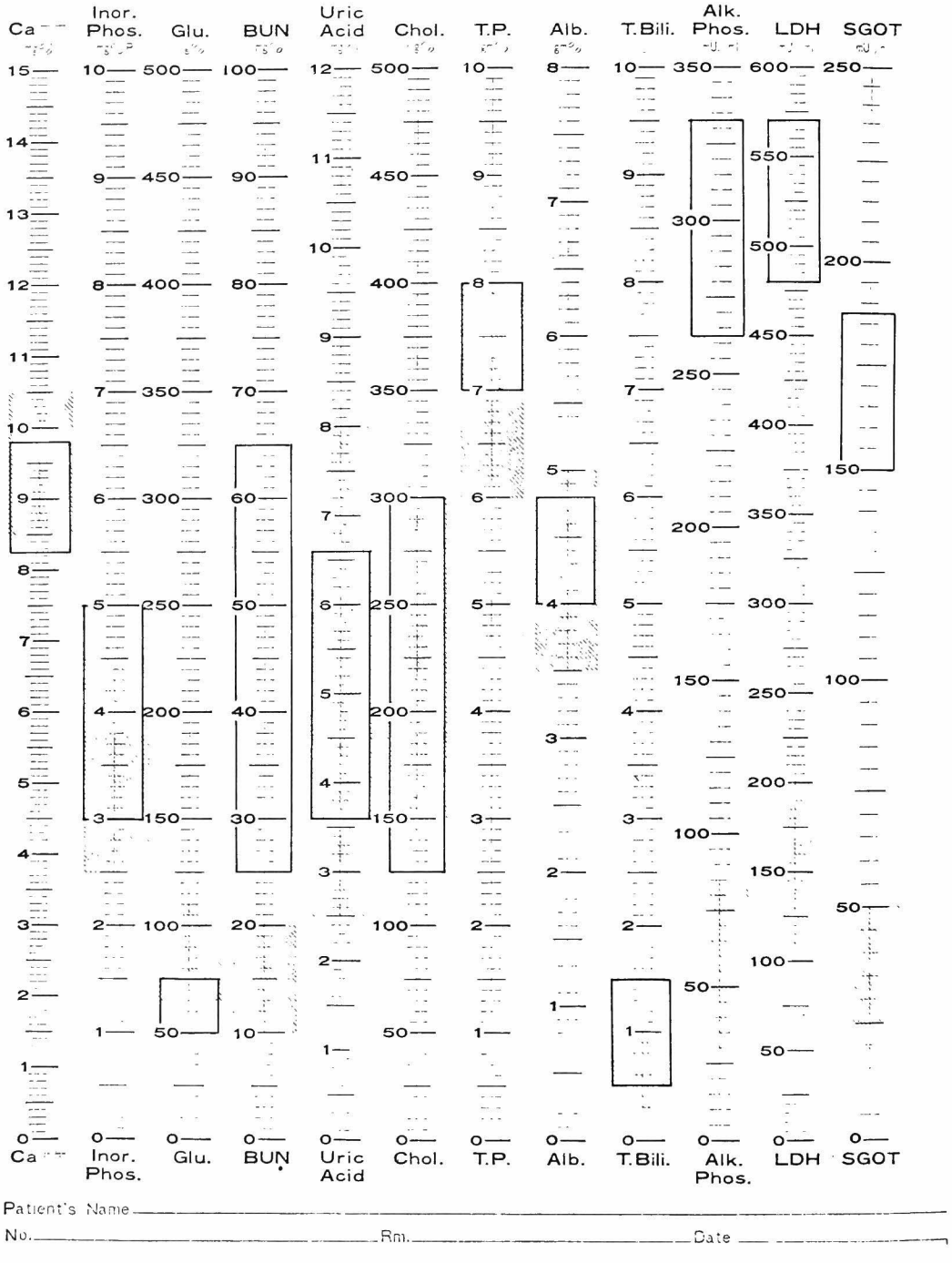


Fig. 3.4.6. Optimum quantizing points(in three-level quantization) for the biochemical data shown by Table 3.2.2, where encircled area in each column represents the interval $[x_1, x_2]$.

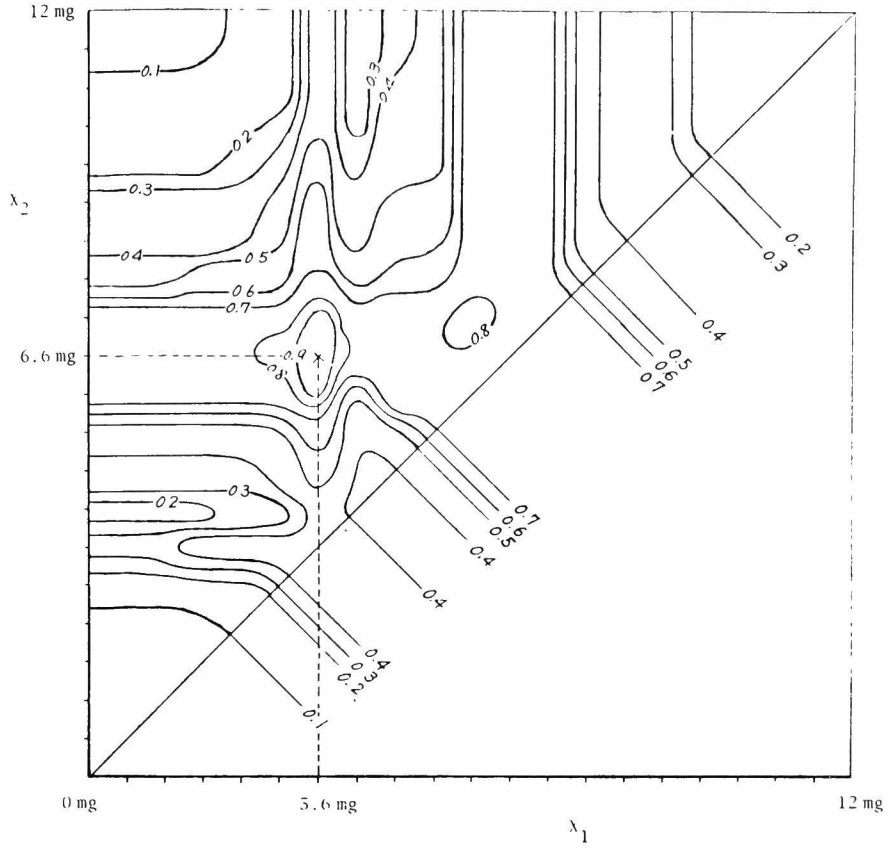


Fig. 3.4.7. The value of mean information intensity for renal failure in uric acid under three-level quantization by points x_1 and x_2, whose maximum is attained by x_1 3.6 mg and x_2 6.6 mg.

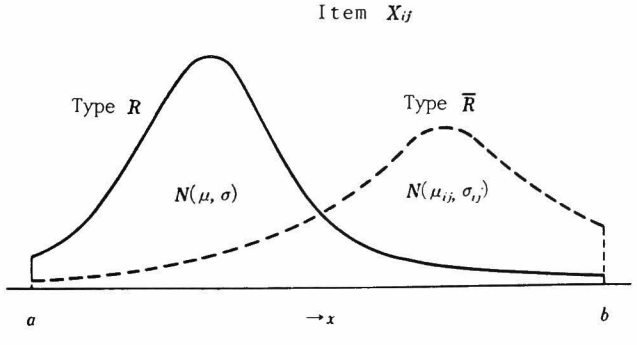


Fig. 3.4.8. Example model for computer simulation.

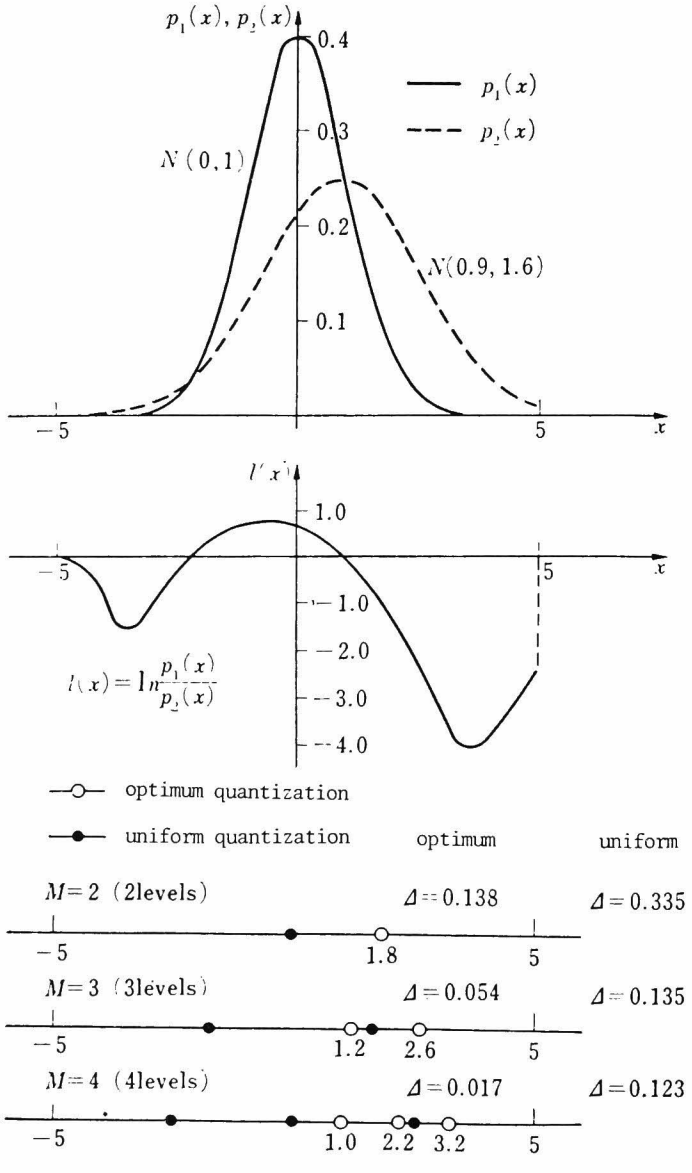


Fig. 3.4.9. Comparisons of measure Δ of information loss by optimum and uniform quantizations and quantizing points for 2, 3, and 4-levels quantization in item X_{43}.

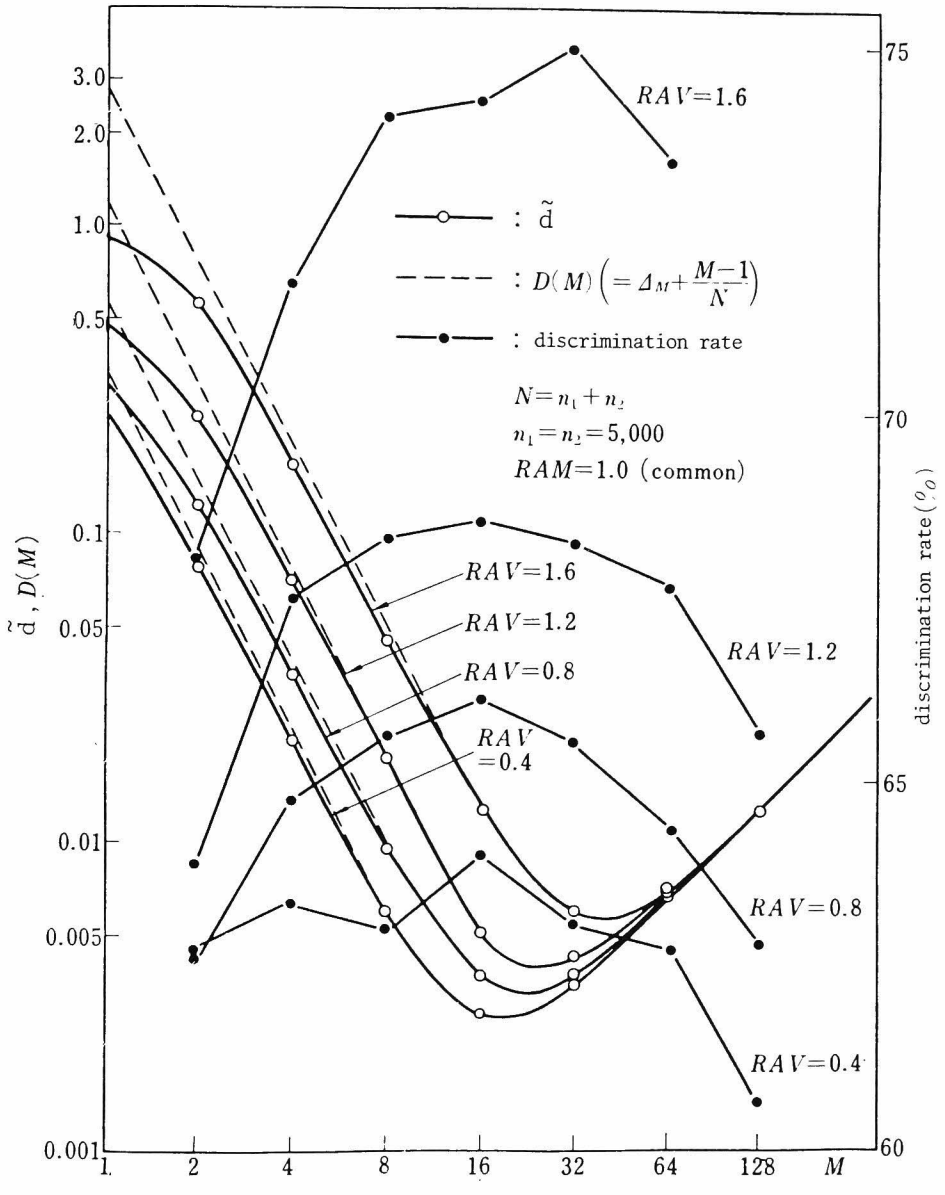


Fig. 3.4.10. Relations among information measures $\tilde{d}$, D(M) and discrimination rate under uniform quantization into 2 to 128 levels.

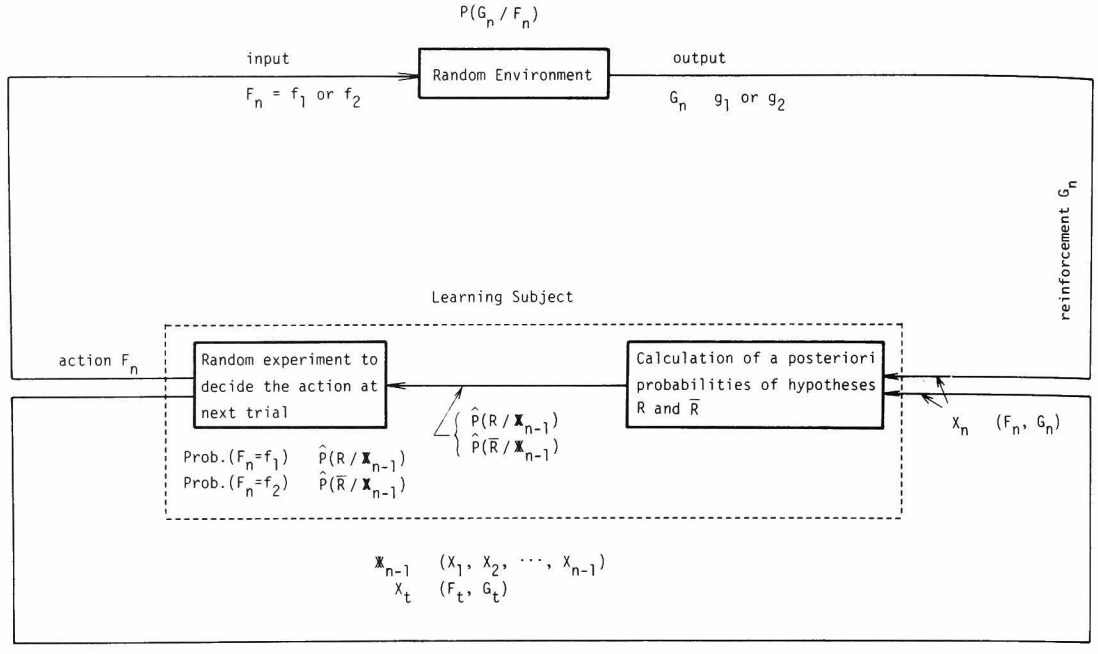


Fig. 3.5.1. Schematic diagram of the learning model.

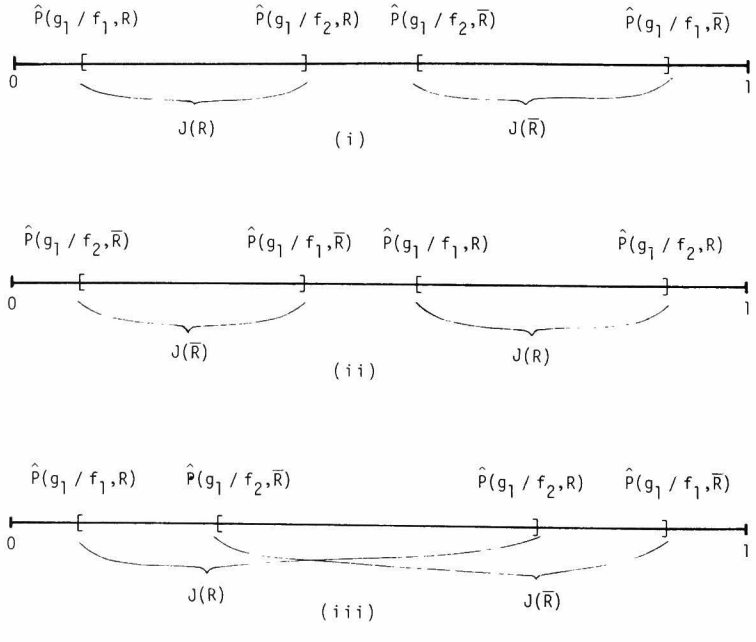


Fig. 3.5.2. Three possible configuration of subjective conditional probabilities $\hat{P}(g_j / f_i, R)$ and $\hat{P}(g_j / f_i, \bar{R})$.

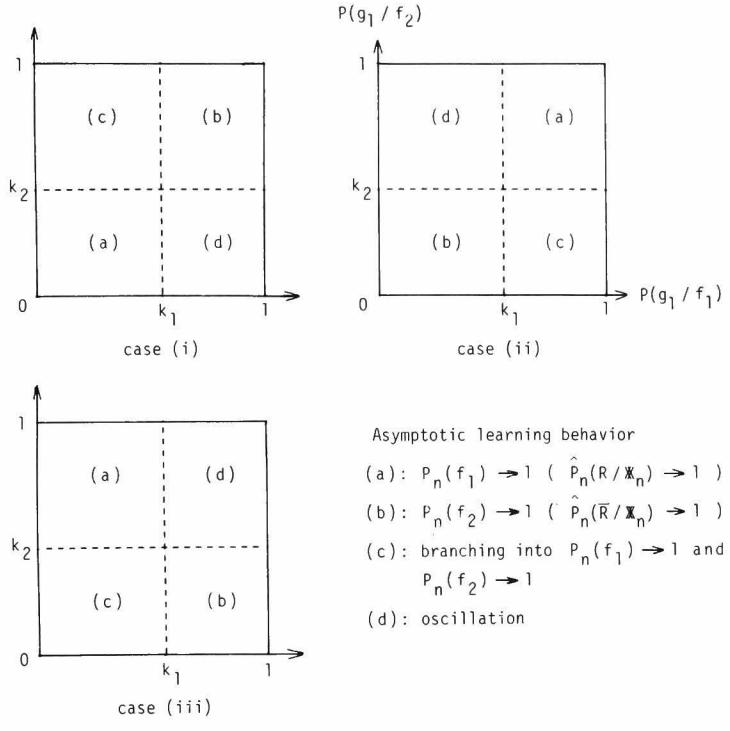


Fig. 3.5.3. Relation of asymptotic learning behavior of the model with objective probabilities $P(g_1 / f_1)$ and $P(g_1 / f_2)$ in case (i), (ii), and (iii) indicated in Fig. 3.5.2.

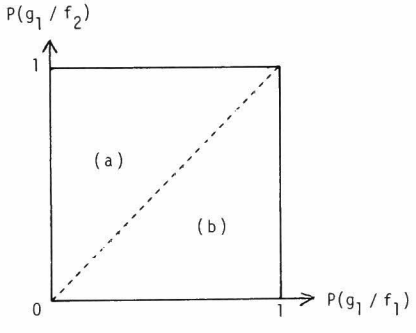


Fig. 3.5.4. The ideal learning behavior suggested by premise for R and $\bar{R}$.

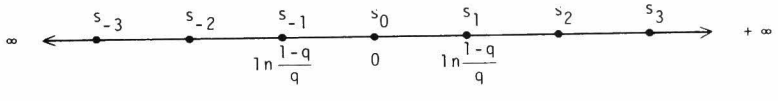


Fig. 3.5.5. Configuration of states of the random process $\{\rho(X_n)\}$, where each interval is equal to $\ln((1-q)/q)$.

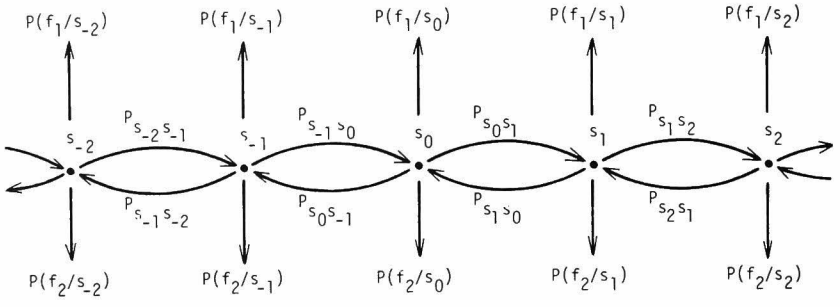


Fig. 3.5.6. Transition diagram of the Markov chain corresponding to the learning model and the response probabilities at each state.

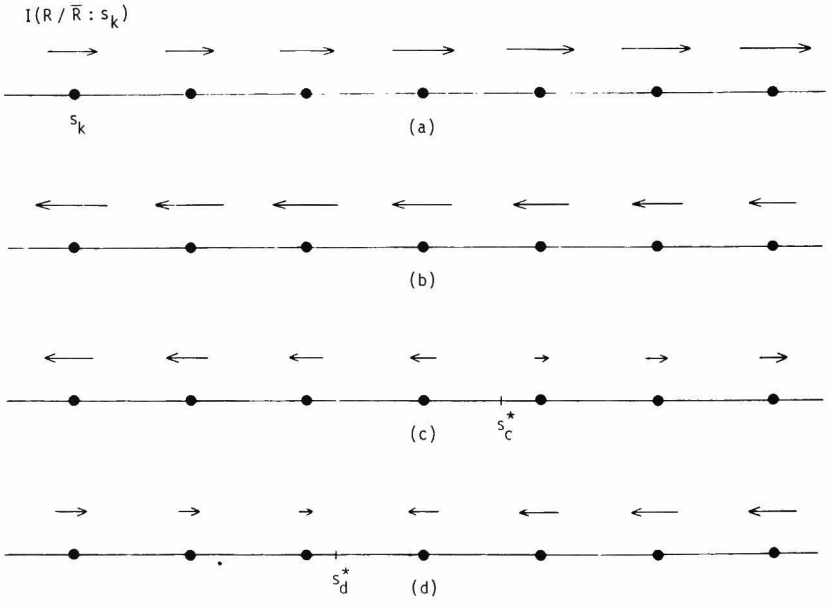


Fig. 3.5.7. Average trend of the transitions in case (a), (b), (c) and (d) of the Norman's classification. Each arrow represents the direction and the magnitude of the average trend $I(R / \bar{R} : s_k)$.

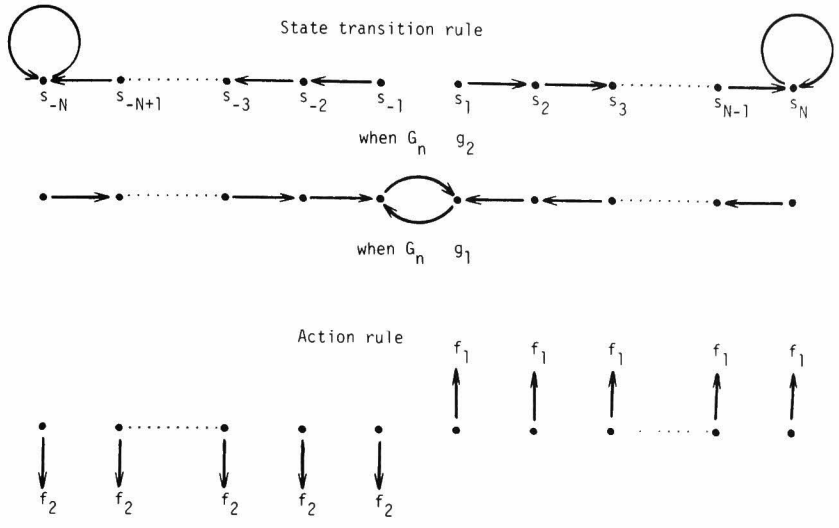


Fig. 3.5.8. State transition and action rule of Tsetlin's learning model called linear tactics.

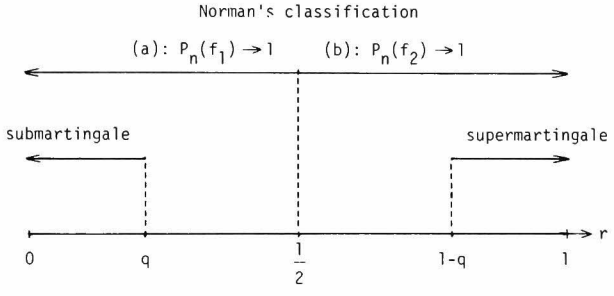


Fig. 3.5.9. Comparison between the Norman's classification and the martingale condition.

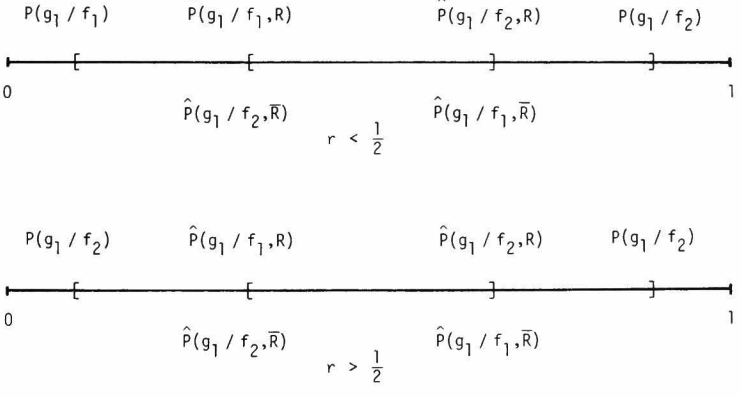


Fig. 3.5.10. Possible configurations of subjective conditional probabilities $\hat{P}(g_j / f_i, R)$ and $\hat{P}(g_j / f_i, \bar{R})$ under the martingale condition.

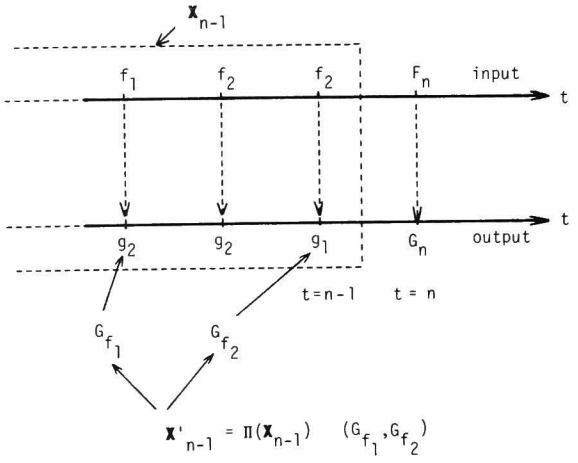


Fig. 3.5.11. An input-output sequence of an environment, where G_{f_i} is the output corresponding to the last f_i input in the past sequence X_{n-1}, for i 1, 2.

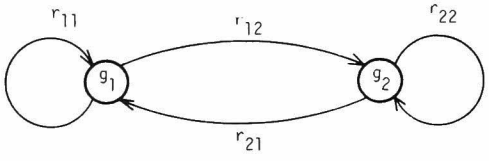
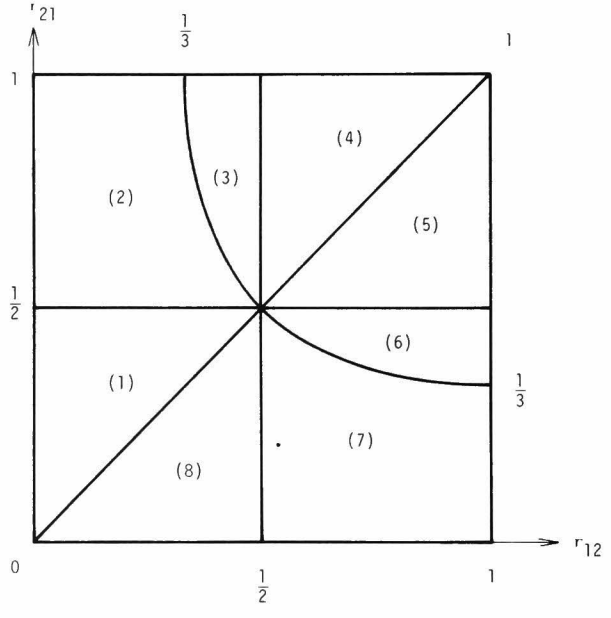


Fig. 3.5.12. Markov chain model of the random environment under action f_1.



(1): I < 0, I' > 0, |I'| > |I|
(2): I < 0, I' > 0, |I'| < |I|
(3): I < 0, I' < 0, |I'| < |I|
(4): I < 0, I' < 0, |I'| > |I|
(5): I > 0, I' < 0, |I'| > |I|
(6): I > 0, I' < 0, |I'| < |I|
(7): I > 0, I' > 0, |I'| < |I|
(8): I > 0, I' > 0, |I'| > |I|

Fig. 3.5.13. The signs of information theoretic measures I and I', and the comparison between their absolute values in relation to the stochastic properties of the environment.

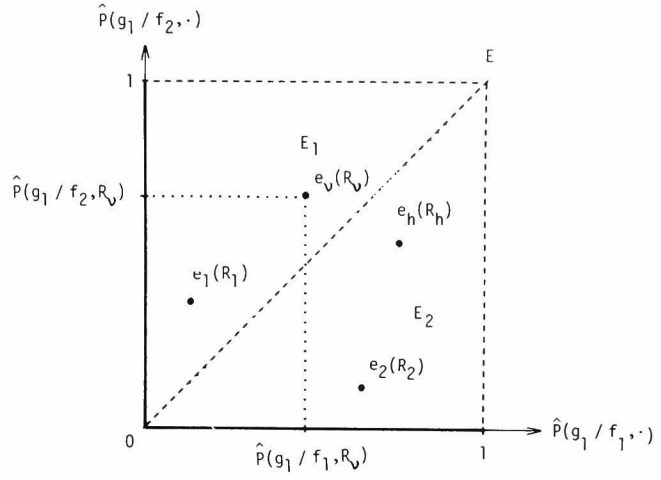


Fig. 3.5.14. Two dimensional representation of multiple hypothses. Each hypothesis R_ν is represented by point $e_\nu = (\hat{P}(g_1 / f_1, R_\nu), \hat{P}(g_1 / f_2, R_\nu))$ in the plane $E = [0, 1] \times [0, 1]$, for ν 1, 2, ···, and h. Regions E_1 and E_2 correspond to actions f_1 and f_2, respectively.

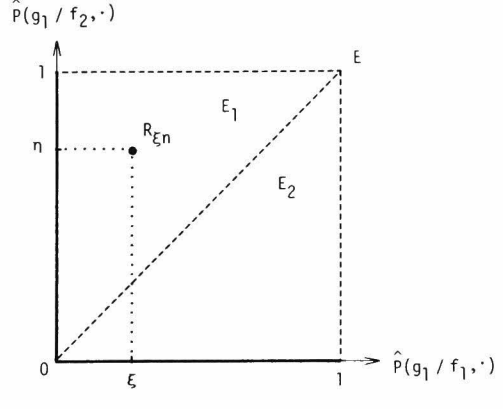


Fig. 3.5.15. Two dimensional representation of a continuum of hypotheses.

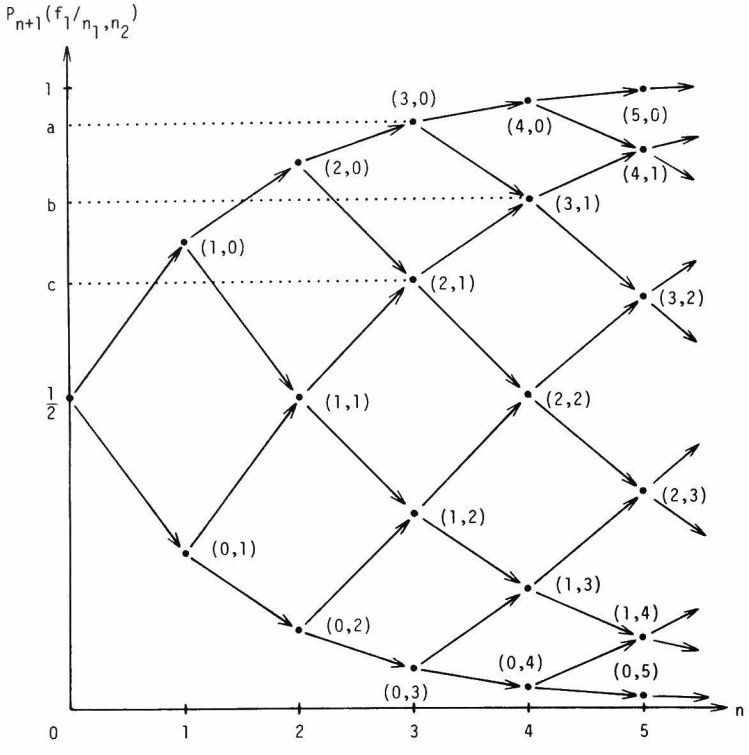


Fig. 3.5.16. Change of response probability $P_{n+1}(f_1 / n_1, n_2)$ of the learning model with a continuum of hypotheses, where each pair of numerals represents (n_1, n_2).

a priori prob.	items / category types / sample types	X_1	X_2	.....	X_n
		$1_1\ 1_2 \cdots 1_{k_1}$	$2_1\ 2_2 \cdots 2_{k_2}$	.....	$n_1\ n_2 \cdots n_{k_n}$
P(R)	R	$p_1^1\ p_2^1 \cdots p_{k_1}^1$	$p_1^2\ p_2^2 \cdots p_{k_2}^2$	.....	$p_1^n\ p_2^n \cdots p_{k_n}^n$
$P(\bar{R})$	$\bar{R}$	$q_1^1\ q_2^2 \cdots q_{k_1}^1$	$q_1^2\ q_2^2 \cdots q_{k_2}^2$	.....	$q_1^n\ q_2^n \cdots q_{k_n}^n$

Table 3.2.1. Response matrix M composed of two sample types R and $\bar{R}$.

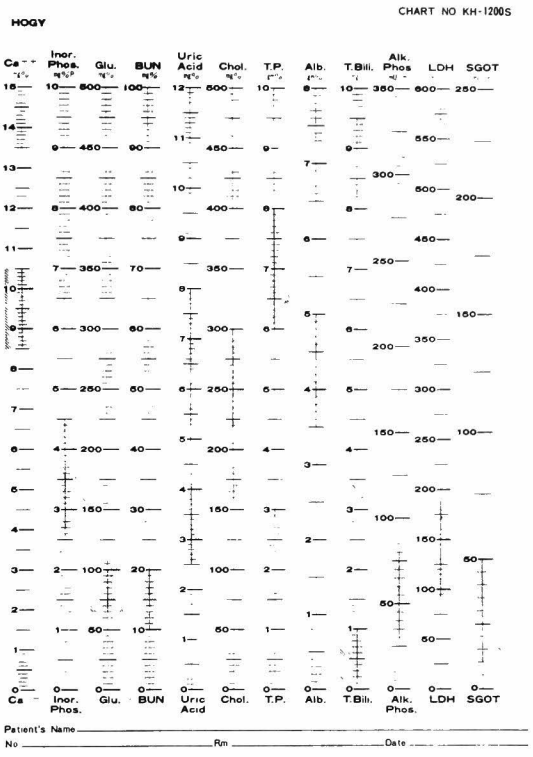


Table 3.2.2. Recording chart for SMA 12/60.

	renal failure		liver failure				diabetes	the other internal diseases			
	neph-rosis	neph-ritis	hepa-titis	cirr-hois	jaun-dice	car-ciroma		hyperpara thyrui-dism	leuke-mia	gout	gestatio-nal foxicosis
Ca	↓	↓→		→↓	↘		↘	↑→			→↓
I.P.	↑→	↑↑					↗				↑→
Glu				↑→			↑↑				
BUN	↑	↑↑									
U.A.	↑	↑↑							↑	↑↑	
Chol.	↑↑	↑	↓	↓	↑↑	↑↓	↑→				
T.P.	↓↓	↓	→↓	→↓	→↓	→↓		↓→			↓
Alb	↓↓	↓	→↓	↓	→↓	→↓			↓		↓
Bil			↑→	↑→	↑↑	↑					
Alk.P.			↑→		↑↑	↑↑		↑			↑
LDH			↑→	→↓		↑↑			↑		
SGOT			↑↑	↑↓	↑→	↑→			↑	↑→	

Table 3.2.3. Physiological characteristics of 12 biochemical blood components in relation to internal diseases.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
R_1	2.722	0.198	0.963	0.290	1.867	2.818	1.515	3.592	0.280	0.588	0.264	0.840
R_2	0.459	0.997	0.004	1.051	0.636	0.174	1.201	1.246	0.387	0.067	0.205	0.001
R_3	0.009	0.036	0.009	0.181	0.098	0.011	0.080	0.027	0.567	0.226	0.031	0.984
R_4	0.382	0.261	1.406	0.089	0.089	0.171	0.211	0.376	0.376	0.073	0.112	0.029
R_5	0.048	0.006	0.017	0.046	0.044	0.113	0.003	0.014	0.017	0.020	0.092	0.009

Table 3.2.4. The value of mean information intensity $S_{R_j}(X_i)$ of each item X_i with respect to each disease R_j for i 1, 2, ···, 12, and j 1, 2, ···, 5.

a priori prob.	sample types	X_1		.....	X_i		.....	X_n	
		1_0	1_1		i_0	i_1		n_0	n_1
P(R)	R	1 - p	p	.....	1 - p	p	.....	1 - p	p
$P(\bar{R})$	$\bar{R}$	p	1 - p	.....	p	1 - p	.....	p	1 - p

Table 3.3.1. Illustrative example of response matrices composed of two sample types.

RAM (=RAV)	discrimination rate P_c	
	theoretical	experimental
0.5	76.6	76.8
1.0	89.5	89.1
1.5	95.3	94.9
2.0	98.0	97.3

Table 3.3.2. Comparison of theoretical and experimental discrimination rates.

	X_i	·····	X_j
	i_1 i_2 ···· i_{k_i}	·····	j_1 j_2 ···· j_{k_j}
R	$P(i_1\|R)$ $P(i_2\|R)$ ··· $P(i_{k_i}\|R)$	·····	$P(j_1\|R)$ $P(j_2\|R)$ ··· $P(j_{k_j}\|R)$
$\bar{R}$	$P(i_1\|\bar{R})$ $P(i_2\|\bar{R})$ ··· $P(i_{k_i}\|\bar{R})$	·····	$P(j_1\|\bar{R})$ $P(j_2\|\bar{R})$ ··· $P(j_{k_j}\|\bar{R})$

(i)

	(X_i, X_j)			
	(i_1,j_1)	(i_1,j_2)	·····	(i_{k_i},j_{k_j})
R	$P(i_1,j_1\|R)$	$P(i_1,j_2\|R)$	·····	$P(i_{k_i},j_{k_j}\|R)$
$\bar{R}$	$P(i_1,j_1\|\bar{R})$	$P(i_1\ j_2\|\bar{R})$	·····	$P(i_{k_i},j_{k_j}\|\bar{R})$

(ii)

Table 3.4.1. Aggregation of two items X_i and X_j into one item (X_i, X_j).

number of quantizing levels M		2	3	4
discrimination rate Pc(%)	optimum	90.1	92.5	92.2
	uniform	77.4	91.2	91.7
information loss Δ_1	optimum	1.74	0.83	0.55
	uniform	3.37	1.50	1.33

Table 3.4.2. Comparison of the optimum and uniform quantizations(for 2, 3, and 4-level quantizations).

item (action) / category (reinforcement) / sample (hypothesis)	f_1		f_2	
	$1_1(g_1)$	$1_2(g_2)$	$2_1(g_1)$	$2_2(g_2)$
R	$P(1_1/R)$	$P(1_2/R)$	$P(2_1/R)$	$P(2_2/R)$
$\bar{R}$	$P(1_1/\bar{R})$	$P(1_2/\bar{R})$	$P(2_1/\bar{R})$	$P(2_2/\bar{R})$

Table 3.5.1. The response table corresponding to a simple learning process, where the sample types, the items and the categories correspond to the hypotheses, the actions and the reinforcements, respectively.

item	X_1				X_2			
category types / sample types	f_1		f_2		f_1		f_2	
	g_1	g_2	g_1	g_2	g_1	g_2	g_1	g_2
R	$P_1(f_1) \times \hat{P}(g_1/f_1,R)$	$P_1(f_1) \times \hat{P}(g_2/f_1,R)$	$P_1(f_2) \times \hat{P}(g_1/f_2,R)$	$P_1(f_2) \times \hat{P}(g_2/f_2,R)$	$P_2(f_1) \times \hat{P}(g_1/f_1,R)$	$P_2(f_1) \times \hat{P}(g_2/f_1,R)$	$P_2(f_2) \times \hat{P}(g_1/f_2,R)$	$P_2(f_2) \times \hat{P}(g_2/f_2,R)$
$\bar{R}$	$P_1(f_1) \times \hat{P}(g_1/f_1,\bar{R})$	$P_1(f_1) \times \hat{P}(g_2/f_1,\bar{R})$	$P_1(f_2) \times \hat{P}(g_1/f_2,\bar{R})$	$P_1(f_2) \times \hat{P}(g_2/f_2,\bar{R})$	$P_2(f_1) \times \hat{P}(g_1/f_1,\bar{R})$	$P_2(f_1) \times \hat{P}(g_2/f_1,\bar{R})$	$P_2(f_2) \times \hat{P}(g_1/f_2,\bar{R})$	$P_2(f_2) \times \hat{P}(g_2/f_2,\bar{R})$

.....	X_n			
.....	f_1		f_2	
.....	g_1	g_2	g_1	g_2
.....	$P_n(f_1) \times P(g_1/f_1,R)$	$P_n(f_1) \times P(g_2/f_1,R)$	$P_n(f_2) \times P(g_1/f_2,R)$	$P_n(f_2) \times P(g_2/f_2,R)$
.....	$P_n(f_1) \times P(g_1/f_1,\bar{R})$	$P_n(f_1) \times P(g_2/f_1,\bar{R})$	$P_n(f_2) \times P(g_1/f_2,\bar{R})$	$P_n(f_2) \times P(g_2/f_2,\bar{R})$

Table 3.5.2. The extended response table at time n, in which each item represents action-reinforcement pairs.

CHAPTER 4 | INTRODUCTION OF A QUANTIFICATION METHOD INTO MARKOV CHAIN STRUCTURES AND ITS APPLICATION TO THE AGGREGATION PROBLEM OF INTER-INDUSTRIAL STRUCTURES

4.1 Introduction

In this chapter, we extend the quantification method III, a method to analyze the internal structure of response matrices through quantifications for the sample types and the category types introduced in Section 2.3.2. This is done in such a way that the extended method is applicable to probabilistic data, i.e., the case where a response matrix represents not merely the existence of sample type-category type relationships but also the frequencies of the relationships.

In Section 4.2, we transform the response matrices into Markov matrices, and the quantification problem is reduced to an analysis of eigenvalue problems of the correlation coefficient between the quantified values for sample types and those for category types. In Section 4.3, the relationship between the above introduced quantifications and the structural properties of the original data is analyzed by introducing the classification of states of Markov chains and also some graph theoretical methods. In Section 4.4, the extended method is applied to the aggregation(or consolidation) problem of Japanese industrial sectors by regarding the input-output table as a response matrix. The aggregation problem is important for forecasting future developments of economy, economical planning, and making global indices of economical activities, and is usually treated by linear algebraic techniques(e.g. refer to Hatanaka[1952] and Ara[1959]).

The problem has a close relation to lumping problems of the states of Markov chains studied by Kemeny & Snell(1963), Thomas & Barr(1977), and Barr & Thomas(1977), et al. These studies are based on linear algebraic methods.

In our method, each industrial sector is quantified in terms of its producing and consuming characteristics and the sectors with nealy the same quantified values are aggregated into one group, hence, our approach is considerably different from the above approaches based on the linear algebraic methods(cf. Katai, Imanaga, & Iwai[1975]).

4.2 Extensions of the Quantification Method III for Probabilistic Data and Investigation of Its Relation to Maximal Correlation Measure

In this section, we consider an extension of the quantification method III introduced in Section 2.3.2 and examine the relation of the extended method to the maximal correlation introduced by Gebelein(1941) and Kramer(1960). In Section 4.2.1, the extension is discussed and the quantification vectors are given as the eigenvectors associated with the maximal eigenvalue of certain Markov matrices. In Section 4.2.2, we discuss the properties of the maximal correlation and the quantification vectors through an extension of the problem discussed in Section 4.2.1.

4.2.1 Extension of the Quantification Method for Probabilistic Data

In this section, we consider the extension of the quatification method III introduced in Section 2.3.2, to allow the response matrix to be composed of not only 0-1 data but also of some real valued positive entities. In other words, the response matrix represents not merely the existence of sample type-category type relationships, but also the strength of the relationships.

Moreover, we confine ourselves to the following special case: the response matrices are consisting of only one item, and the sample types and the category types are composed of the same group of entities, say, 1, 2, $\cdots$, and n, and, hence, the response matrices represent the internal relationships among the corresponding groups. In this case, a response matrix Z can be rewritten as $n \times n$ matrix:

$$Z \quad \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nn} \end{pmatrix}, \tag{4.2.1}$$

where z_{ij} represents the response strength of category type j by sample type i.

We consider the quantification ν_i for the i^{th} sample type and ρ_j for j^{th} category type(i, j = 1, 2, $\cdots$, n) as discussed in Section 2.3.2.

Let us transform(interpret) z_{ij} to the frequency p_{ij} of category type j by sample type i in the following way.

$$p_{ij} = z_{ij} \Big/ \sum_j z_{ij} . \tag{4.2.2}$$

This transformation has no essential effect on quantification vectors $\boldsymbol{\rho}$ and $\boldsymbol{\nu}$.

However, the following interpretation of response matrix becomes possible. The matrix $P = (p_{ij})$ can be considered as a Markov chain with state spase $S = \{1, 2, \cdots, n\}$. The initial distribution vector $\boldsymbol{\pi}$ is given as

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_n), \tag{4.2.3}$$

$$\pi_i = \frac{\sum_j z_{ij}}{\sum_i \sum_j z_{ij}}. \tag{4.2.4}$$

In this case, equations (2.3.28) (2.3.35) in Section 2.3.2 correspond to

$$\tilde{\nu} = \sum_i \nu_i \pi_i = 0, \tag{4.2.5}$$

$$\tilde{\rho} = \sum_j \rho_j \left(\sum_i \pi_i p_{ij}\right) = 0, \tag{4.2.6}$$

$$\sigma_\nu^2 \quad \sum_i \nu_i^2 \pi_i, \tag{4.2.7}$$

$$\sigma_\rho^2 = \sum_j \rho_j^2 \left(\sum_i \pi_i p_{ij}\right), \tag{4.2.8}$$

$$\sigma_{\rho\nu} = \sum_i \sum_j \nu_i \rho_j \pi_i p_{ij}, \tag{4.2.9}$$

$$\kappa_{\rho\nu} \quad \frac{\sigma_{\rho\nu}}{\sigma_\rho \sigma_\nu} = \frac{\sum_i \sum_j \nu_i \rho_j \pi_i p_{ij}}{\sqrt{\sum_i \nu_i^2 \pi_i} \sqrt{\sum_j \rho_j^2 \left(\sum_i \pi_i p_{ij}\right)}}. \tag{4.2.10}$$

Also, the stationary equation (2.3.36) or (2.3.36') corresponds to

$$\left.\begin{aligned} \sum_j \rho_j \pi_i p_{ij} &= \lambda \nu_i \pi_i \\ \sum_i \nu_i \pi_i p_{ij} &= \mu \rho_j \sum_k \pi_k p_{kj} \end{aligned}\right\}, \tag{4.2.11}$$

or

$$\left.\begin{aligned} \sum_j \rho_j p_{ij} &= \lambda \nu_i \\ \sum_i \nu_i \frac{\pi_i p_{ij}}{\sum_k \pi_k p_{kj}} &= \mu \rho_j \end{aligned}\right\}. \tag{4.2.12}$$

In the above, we have assumed

$$\sum_k \pi_k p_{kj} \neq 0 \quad \text{for all } j = 1, 2, \cdots, n, \tag{4.2.13}$$

which is equivqlent to

$$\sum_k z_{kj} \neq 0 \quad \text{for all } j = 1, 2, \cdots, n. \tag{4.2.13'}$$

If $\sum_k z_{kj} \neq 0$, then the j^{th} category type can be removed from the response matrix Z. Hence the above condition can be presumed without loss of generality.

The eigenvalue-eigenvector equation (2.3.37) or (2.3.37') corresponds to

$$\left.\begin{aligned} \sum_k \nu_k \Big(\sum_j p_{ij} \frac{\pi_k p_{kj}}{\sum_h \pi_h p_{hj}} \Big) &= \kappa_{\rho\nu}^2 \nu_i \\ \sum_k \rho_k \Big(\sum_i \frac{\pi_i p_{ij}}{\sum_h \pi_h p_{hj}} p_{ik} \Big) &= \kappa_{\rho\nu}^2 \rho_j \end{aligned}\right\}, \quad \text{for all } i, j = 1, 2, \cdots, n. \tag{4.2.14}$$

When we introduce the next notations:

$$\tilde{P} \triangleq (\tilde{p}_{ij}), \quad \tilde{p}_{ij} \triangleq \frac{\pi_j p_{ij}}{\sum_k \pi_k p_{ki}}, \tag{4.2.15}$$

$$U \triangleq P \tilde{P} = (u_{ij}), \quad u_{ij} = \sum_k p_{ik} \tilde{p}_{kj}, \tag{4.2.16}$$

$$W \triangleq \tilde{P} P = (w_{ij}), \quad w_{ij} = \sum_k \tilde{p}_{ik} p_{kj}, \tag{4.2.17}$$

$$\boldsymbol{\rho} \triangleq \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{pmatrix}, \quad \boldsymbol{\nu} \triangleq \begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{pmatrix}, \tag{4.2.18}$$

then the stationary equation is rewritten as

$$\left.\begin{aligned} P \boldsymbol{\rho} &= \lambda \boldsymbol{\nu} \\ \tilde{P} \boldsymbol{\nu} &= \mu \boldsymbol{\rho} \end{aligned}\right\}. \tag{4.2.19}$$

Also, the eigenvalue-eigenvector equation is equivalent to

$$\left.\begin{aligned} U \boldsymbol{\nu} &= \kappa_{\rho\nu}^2 \boldsymbol{\nu} \\ W \boldsymbol{\rho} &= \kappa_{\rho\nu}^2 \boldsymbol{\rho} \end{aligned}\right\}. \tag{4.2.20}$$

When P is interpreted as a Markov chain, then $\tilde{P}$ can be interpreted as the transition probability of the reversed chain of P with initial probabilities $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_n)$ in the following way(for details, refer to Karlin[1966] and Kemeny & Snell[1963]). Let $\{X_t\}$ be the stationary Markov chain with transition probabilities $P = (p_{ij})$ and initial probability vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_n)$, i.e.,

$$\text{Prob.}(X_0 \quad i) = \pi_i \quad \text{for } i = 1, 2, \cdots, n, \tag{4.2.21}$$

$$\text{Prob.}(X_t \quad j \mid X_{t-1} = i) = p_{ij} \quad \text{for } t \geq 1 \text{ and } i, j = 1, 2, \cdots \cdots, n. \tag{4.2.22}$$

Then, from Bayes theorem, we have

$$\begin{aligned}
&\text{Prob.}(X_{t-1} = j \mid X_t \quad i) \\
&= \frac{\text{Prob.}(X_{t-1} = j,\ X_t \quad i)}{\text{Prob.}(X_t \quad i)} \\
&= \frac{\text{Prob.}(X_{t-1} = j) \cdot \text{Prob.}(X_t \quad i \mid X_{t-1} = j)}{\sum_k \text{Prob.}(X_{t-1} = k) \cdot \text{Prob.}(X_t \quad i \mid X_{t-1} = k)} \\
&\frac{\text{Prob.}(X_{t-1} = j)\ p_{ji}}{\sum_k \text{Prob.}(X_{t-1} = k)\ p_{ki}} .
\end{aligned} \tag{4.2.23}$$

Particularly, when we set t = 1, then the above equation yields

$$\text{Prob.}(X_0 = j \mid X_1 = i) = \frac{\pi_j p_{ji}}{\sum_k \pi_k p_{ki}}$$

$$\tilde{p}_{ij} \quad \text{for } i, j = 1, 2, \cdots, n. \tag{4.2.24}$$

However, the reversed chain has not necessarily the Markov Property(cf. Burke & Rosenblatt[1958]). It is a Markov chain only when $\text{Prob.}(X_{t-1} \quad i)$ does not depend on t for i = 1, 2, ···, n, i.e., only when the initial probalility vector $\boldsymbol{\pi}$ coincides with the stationary probability vector $\mathbf{a}$;

$$\boldsymbol{\pi} = \mathbf{a}. \tag{4.2.25}$$

We have shown that $\tilde{P}$ is a Markov matrix, hence, U $P\tilde{P}$ and $W = \tilde{P}P$ are also Markov matrices. Furthermore, as shown in the next section, U and W are positive definite matrices. Therefore, the eigenvalues of U and W are nonnegative real values and and are equal to or less than 1. The eigenvectors can be set as being real valued. From the definition that

$$u_{ij} = \sum_k p_{ik} p_{jk} \pi_j \Big/ \sum_h \pi_h p_{hk} \quad \text{for } i, j = 1, 2, \cdots, n, \tag{4.2.16'}$$

the quantity u_{ij} can be considered as the degree of the similarity between the response pattern of the i^{th} sample type that of the j^{th} sample type.

Also, from

$$w_{ij} \quad \sum_k p_{ki}\, p_{kj}\, \pi_k \;/\; \sum_h \pi_h\, p_{hi} \quad \text{for } i, j = 1, 2, \cdots, n, \qquad (4.2.17')$$

w_{ij} can be regarded as the degree of proximity between the i^{th} category type and j^{th} category type in the sense that both of them are outcomes from the same sample type.

In the original notation z_{ij}, they are given as

$$u_{ij} = \sum_k \left(\frac{z_{ik}}{\sum_h z_{ih}} \frac{z_{jk}}{\sum_h z_{hk}} \right), \qquad (4.2.26)$$

$$w_{ij} = \sum_k \left(\frac{z_{ki}}{\sum_h z_{hi}} \frac{z_{kj}}{\sum_h z_{kh}} \right). \qquad (4.2.27)$$

4.2.2 Extensions of the Method to the Case of Continuum of Sample Types and Category Types and Investigation of Its Relation to Maximal Correlation Measure

The discussions in the previous section can be extended to a more general case where ranges of the sample types and the category types are given as a continuum, respectively. This extended discussion is mainly based on the studies of the evaluation of the dependence between two random variables made by Gebelein(1941), Kramer(1960), and Rényi(1959b).

Let s and c denote two random variables corresponding to the sample type and category type, respectively. Also, let p(s,c) be the joint probability distribution function and the marginal distribution be as as follows:

$$p(s) = \int p(s,c)\, dc, \qquad (4.2.28)$$

$$p(c) = \int p(s,c)\, ds. \qquad (4.2.29)$$

Let us consider $\nu(s)$, the quantification of s, and the quantification of c, $\rho(c)$, such that the correlation coefficient $\kappa_{\rho\nu}$ of the random variables $\nu(s)$ and $\rho(c)$ be the maximum, where

$$\kappa_{\rho\nu} = \frac{\sigma_{\rho\nu}}{\sigma_\rho\, \sigma_\nu}, \qquad (4.2.30)$$

$$\sigma_\nu^2 = \int\int \nu^2(s)\, p(s,c)\, ds\, dc \qquad \int \nu^2(s)\, p(s)\, ds, \qquad (4.2.31)$$

$$\sigma_\rho^2 = \int\int \rho^2(c)\, p(s,c)\, ds\, dc \quad \int \rho^2(c)\, p(c)\, dc, \tag{4.2.32}$$

$$\sigma_{\rho\nu} = \int\int \nu(s)\, \rho(c)\, p(s,c)\, ds\, dc. \tag{4.2.33}$$

For simplicity, we set ρ and ν as

$$\sigma_\rho^2 = \sigma_\nu^2 = 1. \tag{4.2.34}$$

Then, ρ and ν are given by the following variational problem(cf. Gebelein[1941]):

$$\kappa_{\rho\nu}^2 = \{ \int\int \nu(s)\, \rho(c)\, p(s,c)\, ds\, dc \}^2 \longrightarrow \max. \tag{4.2.35}$$

To solve the above problem, the following notations are introduced.

$$N(s) \triangleq \nu(s) \sqrt{p(s)}, \tag{4.2.36}$$

$$R(s) \triangleq \frac{1}{\sqrt{p(s)}} \int \rho(c)\, p(s,c)\, dc. \tag{4.2.37}$$

Then $\kappa_{\rho\nu}^2$ becomes

$$\kappa_{\rho\nu}^2 = \left(\int N(s)\, R(s)\, ds \right)^2 \tag{4.2.38}$$

From schwarz inequality and (4.2.33),

$$\begin{aligned} \kappa_{\rho\nu}^2 &\le \int N^2(s)\, ds \cdot \int R^2(s)\, ds \\ &= \int N^2(s)\, ds \\ &= \int \frac{1}{p(s)} \left(\int \rho(c)\, p(s,c)\, dc \right)^2 ds, \end{aligned} \tag{4.2.39}$$

where the equality holds only when

$$\frac{R(s)}{N(s)} = \frac{\int \rho(c)\, p(s,c)\, dc}{\nu(s)\, p(s)} \quad k(\text{const.}). \tag{4.2.40}$$

In this case, we have

$$\begin{aligned} k^2 &= \int R^2(s)\, ds \\ &= \int \frac{1}{p(s)} \left(\int \rho(c)\, p(s,c)\, dc \right)^2 ds \\ &= (\kappa_{\rho\nu})_{\max}^2, \end{aligned} \tag{4.2.41}$$

$$\nu(s) \quad \frac{1}{k\ p(s)} \int \rho(c)\ p(s,c)\ dc. \tag{4.2.42}$$

Due to the symmetry of criterion (4.2.35) with respect to $\nu(s)$ and $\rho(c)$, we also obtain

$$\rho(c) = \frac{1}{k'\ p(c)} \int \nu(s)\ p(s,c)\ ds. \tag{4.2.43}$$

From (4.2.41), (4.2.42), and (4.2.43), the variational problem (4.2.35) is reduced to

$$\nu(s) = \frac{1}{(\kappa_{\rho\nu})_{max}} \int \frac{p(s,c)}{p(s)} \rho(c)\ dc, \tag{4.2.44}$$

$$\rho(c) = \frac{1}{(\kappa_{\rho\nu})_{max}} \int \frac{p(s,c)}{p(c)} \nu(s)\ ds. \tag{4.2.45}$$

From the above two equations, we obtain

$$\nu(s) = \frac{1}{(\kappa_{\rho\nu})_{max}^{2}} \int \frac{P(s,s')}{p(s)} \nu(s')\ ds', \tag{4.2.46}$$

where

$$P(s,s') \triangleq \int \frac{p(s,c)\ p(s',c)}{p(c)}\ dc\ . \tag{4.2.47}$$

Or

$$\int Q(s,s')\ N(s')\ ds' \quad (\kappa_{\rho\nu})_{max}^{2}\ N(s) = 0, \tag{4.2.48}$$

where

$$Q(s,s') \triangleq \int \frac{p(s,c)\ p(s',c)}{\sqrt{p(s)\ p(s')}} \frac{1}{p(c)}\ dc \tag{4.2.49}$$

and $N(s)$ is given by (4.2:35).

Equations (4.2.46) and (4.2.48) are Fredholm type integral equations, and $(\kappa_{\rho\nu})_{max}^{2}$ is its eigenvalue, and $\nu(s)$ (or $\rho(c)$) is its eigenfunction. In (4.2.48), $Q(s,s')$ is a symmetric and positive definite kernel, hence there must exist eigenvalues and eigenfunctions, and all of them are real valued. It is easy to see that $\nu(s) \equiv 1$, $\rho(c) \equiv 1$, and $(\kappa_{\rho\nu})_{max} = 1$ are a solution of the

above integral equation, however, they are meaningless from the standpoint of the quantification methods. Hence we supplement the next condition on ν and ρ.

$$\int \nu(s)\, p(s)\, ds = \int\int \nu(s)\, p(s,c)\, ds\, dc = 0. \tag{4.2.50}$$

$$\int \rho(c)\, p(c)\, dc = \int\int \rho(c)\, p(s,c)\, ds\, dc = 0. \tag{4.2.51}$$

The above condition only rejects(omits) the meanigless case: $\nu(s) \equiv \rho(c) \equiv 1$.

Then the variational problem (4.2.35) and the integral equations (4.2.26) and (4.2.48) can be rewritten respectively as

$$\kappa_{\rho\nu}^2 = \int\int \left(P(s,s') - p(s)\, p(s') \right) \nu(s)\, \nu(s')\, ds\, ds' \longrightarrow \max. \tag{4.2.35'}$$

$$\nu(s) = \frac{1}{(\kappa_{\rho\nu})_{\max}^2} \int \frac{P(s,s') - p(s)\, p(s')}{p(s)}\, \nu(s')\, ds \tag{4.2.46'}$$

or

$$\int Q'(s,s')\, N(s')\, ds' - (\kappa_{\rho\nu})_{\max}^2\, N(s) = 0, \tag{4.2.48'}$$

where

$$Q'(s,s') \triangleq \int \frac{p(s,c)\, p(s',c) - p(s)\, p(s')\, p(c)^2}{p(c)\sqrt{p(s)\, p(s')}}\, dc. \tag{4.2.49'}$$

Let $\kappa_0^2, \kappa_1^2, \cdots$ be the eigenvalues of integral equation (4.2.48) such that

$$\kappa_0 = 1 \geq \kappa_1^2 \geq \kappa_2^2 \geq \cdots \tag{4.2.52}$$

Also, let $N_i(s) = \nu_i(s)\sqrt{p(s)}$ be the associated eigenfunction for $i = 1, 2, \cdots$. Then, because of $Q(s,s')$ being a symmetric positive definite kernel, we can apply Mercer's theorem, which yields

$$Q(s,s') = \sum_{i=0}^{\infty} \kappa_i^2\, N_i(s)\, N_i(s')$$

$$\sum_{i=0}^{\infty} \kappa_i^2 \sqrt{p(s)\, p(s')}\, \nu_i(s)\, \nu_i(s'). \tag{4.2.53}$$

Integrating the above equation for $s = s'$, we have

$$\int Q(s,s)\, ds = \sum_{i=0}^{\infty} \kappa_i^2. \tag{4.2.54}$$

Let χ^2 be the above sum minus $\kappa_0^2 = 1$, i.e.,

$$\chi^2 \triangleq \sum_{i=1}^{\infty} \kappa_i^2$$

$$\int Q(s,s')\,ds - 1$$

$$= \int\int \frac{p(s,c)^2}{p(s)\,p(c)} - \frac{p(s)^2\,p(c)^2}{p(s)\,p(c)}\,ds\,dc$$

$$= \int\int \frac{(p(s,c) - p(s)\,p(c))^2}{p(s)\,p(c)}\,ds\,dc. \qquad (4.2.55)$$

The above value corresponds to a kind of statistical measure , mean square contingency, which evaluates the degree of dependence between the random variables s and c. The measure is used in statistical testing whether s and c are stochastically independent or not in the following way, where s and c are finitely valued, i.e., $s = s_1, s_2, \cdots, s_u$ and $c \quad c_1, c_2, \cdots, c_v$. Let $p(s_i)$, $p(c_j)$, $p(s_i,c_j)$ be the frequency of s_i, c_j, and (s_i,c_j) from a set of data. Then calculate

$$\chi^2 \quad \sum_{i=1}^{u} \sum_{j=1}^{v} \frac{(p(s_i,c_j) - p(s_i)\,p(c_j))^2}{p(s_i)\,p(c_j)}. \qquad (4.2.56)$$

If χ^2 is larger than certain critical value, then s and c are regarded as not being independent.

In the above,

$$(\kappa_{\rho\nu})_{max} \quad \kappa_1, \qquad (4.2.57)$$

hence

$$\chi^2 \quad (\kappa_{\rho\nu})_{max}^2 + \kappa_2^2 + \cdots. \qquad (4.2.58)$$

Gebelein(1941) and Kramer(1960) called $\kappa_1 = (\kappa_{\rho\nu})_{max}$ *maximal correlation* between the random variables s and c. For the measure $\delta(s,c)$ of dependence between two random variables s and c, Rényi(1956b) gave the following postulates:

(1) $\delta(s,c)$ is defined for any pair of random variables s and c, neither of them being constant with probability 1.

(2) $\delta(s,c) \quad \delta(c,s)$.

(3) $0 \le \delta(s,c) \le 1$.

(4) $\delta(s,c) = 0$ iff s and c are stochastically independent.

(5) $\delta(s,c) = 1$ if there exists dependence between s and c, i.e., either s $g(c)$ or $c = h(s)$, where g and h are Borel-measurable functions.

(6) If Borel-measurable functions $u(x)$ and $v(x)$ map the real axis in a one-to-one manner onto itself, then $\delta(u(s),v(c)) = \delta(s,c)$.

(7) If the joint distribution of s and c is normal, i.e., (s,c) is Gaussian, then $\delta(s,c) = |K(s,c)|$, where $K(s,c)$ is the correlation coefficient of s and c.

He has shown that among the dependence measures introduced so far, only the maximal correlation and Linfoot's information theoretic measure(cf. Linfoot [1959] and Hamdan & Tsokos[1971]):

$$L(s,c) = \sqrt{1 \quad \exp(\ 2\ M(s,c))} \tag{4.2.59}$$

satisfies all the above seven postulates, where $M(s,c)$ is the Shannon's mutual information between s and c. Rényi(1959a) also applied the measure, maximal correlation, to solve some mathematical problems in number theory. Bell(1958 & 1962) examined the measure in connection with information theory.

The response probability (4.2.2) corresponds to

$$p(c|s) \triangleq \frac{p(s,c)}{p(s)} . \tag{4.2.60}$$

The relation of the above response probability to the quantifications $\rho_i(c)$, $\nu_i(s)$, and the eigenvalues $\kappa_i(i \quad 1, 2, \cdots)$ is summarized by the following theorem(cf. Katai, Imanaga, & Iwai[1975]):

THEOREM 4.2.1 :

$$p(c|s) = p(c)\cdot\{ \sum_{i=0}^{\infty} \kappa_i\ \nu_i(s)\ \rho_i(c)\ \}$$

$$= p(c)\cdot\{\ 1 + (\kappa_{\rho\nu})_{max}\ \nu_1(s)\ \rho_1(c) + \kappa_2\ \nu_2(s)\ \rho_2(c) + \cdots\ \} \tag{4.2.61}$$

$$\int (\ p(c|s) \quad p(c|s')\ )^2 / p(c)\ dc \quad \sum_{i=1}^{\infty} \kappa_i^{\,2}\ (\ \nu_i(s) \quad \nu_i(s')\ )^2$$

$$= (\kappa_{\rho\nu})_{max}^{\,2}\ (\ \nu_1(s) \quad \nu_1(s')\ )^2$$

$$+ \sum_{i=2}^{\infty} \kappa_i^{\,2}\ (\ \nu_i(s) \quad \nu_i(s')\ )^2 . \tag{4.2.62}$$

PROOF : From $Q(s,s')$ being symmetric positive definite kernel, the set of eigenfunctions $\{N_i(s)\}$ $\{\sqrt{p(s)}\ \nu_i(s)\}$ constitutes a complete orthonormal system. Also, from (4.2.45), we have

$$\sqrt{p(c)}\ \rho_i(c) = \frac{1}{\kappa_i}\int \frac{p(s,c)}{\sqrt{p(c)}\ \sqrt{p(s)}}\ \sqrt{p(s)}\ \nu_i(s)\ ds, \tag{4.2.63}$$

it follows that

$$\frac{p(s,c)}{\sqrt{p(c)}\ \sqrt{p(s)}} = \sum_{i=0}^{\infty} \kappa_i \sqrt{p(c)}\ \rho_i(c)\ \sqrt{p(s)}\ \nu_i(s). \tag{4.2.64}$$

Hence we obtain

$$p(c|s) = \frac{p(s,c)}{p(s)} \quad p(c)\cdot\{\sum_{i=0}^{\infty} \kappa_i\ \nu_i(s)\ \rho_i(c)\}. \tag{4.2.65}$$

Using the similar method, it is easy to show that $\{\sqrt{p(c)}\ \rho_i(c)\}$ also constitutes a complete orthonormal system. Therefore, from (4.2.65), we have

$$\int (p(c|s) - p(c|s'))^2 / p(c)\ dc = \sum_{i=0}^{\infty} \kappa_i^{\,2}\ (\ \nu_i(s) - \nu_i(s')\ )^2. \tag{4.2.66}$$

Because of $\nu_0(s) \equiv 1$, we finally obtain (4.2.62). □

From the above theorem, we can say that the first "mode" of the difference between $p(c|s)$ and $p(c)$, the effect of sample type s to the probability $p(c)$, is given as $(\kappa_{\rho\nu})_{max}\nu_1(s)\rho_1(c)$, i.e., the first mode is determined by the quantification for sample type $\nu_1(s)$ and the quantification for category type $\rho_1(c)$, and the maximal correlation $(\kappa_{\rho\nu})_{max}$. Also, from (4.2.62), the first mode of the difference is given as $(\kappa_{\rho\nu})_{max}(\nu_1(s) \quad \nu_1(s'))^2$, i.e., the first mode is given as $(\nu_1(s) \quad \nu_1(s'))^2$, the square difference between the quantification $\nu_1(s)$ and $\nu_1(s')$, times $(\kappa_{\rho\nu})_{max}^{\,2}$, the square of the maximal correlation

In Section 4.2, it was clarified that the extended problem of the quantification method III was also reduced to an eigenvalue problem of Markov matrices U and W as shown in (4.2.20). The meaning of the matrices U and W(cf. (4.2.16) & (4.2.17)) was analyzed and also the properties of correlation coefficient $(\kappa_{\rho\nu})_{max}$ between quantification ν for sample types and quantification ρ for category types were discussed. The coefficient $(\kappa_{\rho\nu})_{max}$ has desirable properties as a measure of dependence between sample types and category types. Moreover, through an extension of the problem, we have clarified the dependence of response probability p_{ij} to the quantification vectors $\boldsymbol{\nu}$, $\boldsymbol{\rho}$, and the coefficient $(\kappa_{\rho\nu})_{max}$(cf. theorem 4.2.1).

4.3 Qualitative Properties of the Quantification Method

In the preceding section, we introduced the quantification vector $\boldsymbol{\nu}$, which is defined as the eigenvector of Markov matrix U, by transforming the response matrix M into the Markov matrix P. Hence the quantification method is closely

related to the structural properties of Markov chains.

In Section 4.3.1, we refer to the classification of states of Markov chains and also to the classification of Markov chains themselves. In Section 4.3.2, we discuss the algebraic properties of Markov chains: the relationship between the above classifications and the eigenvectors of Markov matrices. In Section 4.3.3, we consider the graph theoretical descriptions of Markov chains and also of the classifications in Section 4.3.1. Particularly, the relationship between the classification of states in the Markov chain given by U and that in the Markov chains given by P is discussed.

4.3.1 Classification of the States of Markov Chains and Classification of the Types of Markov Chains

Let $\{X_t\}$ be a stationary Markov chain and its state space S be $\{1, 2, \cdots, n\}$. Then the transition probability p_{ij} is represented as

$$p_{ij} \quad \text{Prob.}(X_{t+1} = j \mid X_t \quad i) \quad \text{for } t \geq 0. \tag{4.3.1}$$

We denote the initial distribution by vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_n)$ and the stationary distribution(invariant distribution) by $\mathbf{a}$ $(a_1, a_2, \cdots, a_n)$. The following integer h_{ij} is called (the first) hitting time for state j from state i.

$$h_{ij} = \min.[t \geq 1 \mid X_0 = i, X_t \quad j]. \tag{4.3.2}$$

Of course, the integer h_{ij} is a random variable and gives the next classification of states(cf. Kemeny & Snell[1963] and Romanovsky[1970]).

If $\text{Prob.}(h_{ii} < \infty) = 1$, then state i is called recurrent. If $\text{Prob.}(h_{ii} < \infty) < 1$, then it is called transient. (4.3.3)

The set of the recurrent states, i.e., $\{i \mid \text{Prob.}(h_{ii} < \infty) = 1\}$ is called recurrent class and the set of transient states, i.e. $\{i \mid \text{Prob.}(h_{ii} < \infty) < 1\}$ is called transient class.

The notion of hitting time also gives the relationships between states in the following way: If $\text{Prob.}(h_{ij} < \infty) > 0$, then we write $i \to j$ and say that j is reachable from i(in Section 4.3.3, we will give its graph theoretical description). If $\text{Prob.}(h_{ij} < \infty) \quad 0$, then we write $i \not\to j$(j is not reachable from i)(cf. Kemeny & Snell[1963] and Iwahori[1974]).

According to the binary relation "$\to$", the above classification of states is interpreted as

State i is transient iff there exists a state $j(\neq i)$ with $i \to j$ and $j \to i$. State i is recurrent iff $j \to i$ for any j such that $i \to j$. (4.3.4)

When $i \to j$ and $j \to i$, we write $i \leftrightarrow j$. Let R and T be the recurrent class and the transient class of state space S, respectively. Then, for the states in R, the relation " $\leftrightarrow$ " is an equivalence relation. R can be partitioned into equivalence classes $\{E_r\}$, and each equivalence class E_r is called an ergodic class. An ergodic class consisting of a single state is called an absorbing state, i.e., state j is absorbing iff $j \not\to k$ for all $k \neq j$. For an ergodic class E, it can be easily shown that the greatest common divisor d of $\{t > 1 \mid \text{Prob.}(X_0 = i \text{ and } X_t = i) > 0\}$ for $i \in E$ does not depend on the selection of $i \in E$. The number d is called the period of class E, and E is called a cyclic class with period d, when $d > 1$. A cyclic class E with period d can be decomposed into d moving classes(cyclic parts) $F_1, F_2, \cdots$, and F_d in the following way:

$$F_t = \{j \in E \mid \text{Prob.}(X_t = j \mid X_0 = i) > 0\}$$

for fixed i and $t = 1, 2, \cdots, d$. (4.3.5)

The moving classes have the properties:

$$F_t \cap F_{t'} = \phi \quad (\text{for } t \neq t'),$$

$$\#F_t = \#E / d,$$

$$\text{Prob.}(X_{t+1} \in F_{t+1} \mid X_t = i) = 1 \text{ for any } i \in F_t \text{ and any } t \geq 0. \quad (4.3.6)$$

An ergodic class with period 1 is called regular.

Using the above classification of states, Markov chains themselves can be classified into the following types:

[1] Chains without transient states.

a. The number of ergodic class is one.

a.1. The ergodic class is regular ····· "ergodic chain".

a.2. The ergodic class is cyclic ····· "cyclic chain".

b. The number of ergodic classes is more than one.

[2] Chains with transient states.

a. All the ergodic classes are regular.

a.1. Each ergodic class is consisted of a single state ··· "absorbing chain".

a.2. An ergodic class contains more than one state.

b. All the ergodic classes are cyclic.

c. They have both regular and cyclic classes.

4.3.2 Algebraic Properties of Markov Chains

In this section , we consider the relations between the eigenvalues of transition probability matrices and the classifications of states or chains introduced in the previous section(cf. Karlin[1966] and Romanovsky[1970]).

It will be easily seen that the eigenvalues λ's of a Markov matrix satisfies

$$|\lambda| \leq 1. \tag{4.3.7}$$

On the other hand, the vector ${}^t(1, 1, \cdots, 1)$ is immediately seen to be a right eigenvector of P associated with eigenvalue 1; thus the spectral redius of P is equal to 1, i.e., $\max |\lambda| = 1$. The left eigenvector of P assiciated with eigenvalue 1 is $\mathbf{a}$, the stationary distribution vector.

Relabeling the states, if necessary, we may assume that

$$R = \{1, 2, \cdots, r\},$$

$$T = \{r+1, r+2, \cdots, n\}. \tag{4.3.8}$$

From the definition of recurrent class R and transient class T, we have

$$p_{ij} \quad 0 \quad \text{for any pair of i and j s.t. } i \in R \text{ and } j \in T. \tag{4.3.9}$$

Thus P has the form

$$P = \begin{pmatrix} P' & 0 \\ B & C \end{pmatrix}, \tag{4.3.10}$$

and P' forms an $r \times r$ Markov matrix. Noting that the vector $\mathbf{a}$ is also a left eigenvector of P^k for any $k \geq 1$ and the form indicated by (4.3.10), we obtain

$$a_{r+1} = a_{r+2} \quad \cdots \quad a_n = 0, \text{ i.e.,}$$

$$\mathbf{a} = (a_1, a_2, \cdots, a_r, 0, 0, \cdots, 0). \tag{4.3.11}$$

The vector $\mathbf{a}' = (a_1, a_2, \cdots, a_r)$ is a left eigenvector of P'.

Suppose that R is composed of h ergodic classes E_1, E_2, $\cdots$, and E_h. Then, relabelling the states if necessary, P has the following form.

$$P = \begin{pmatrix} P_1 & & & 0 \\ & P_2 & 0 & 0 \\ & & \ddots & \\ & 0 & & P_h \; 0 \\ B_1 & B_2 & \cdots & B_h \; C \end{pmatrix}. \tag{4.3.12}$$

Hence

$$\begin{aligned} \mathbf{a}^{(1)} &= (a_1, a_2, \cdots, a_{n_1}, 0, 0, \cdots \quad \cdots, 0) \\ \mathbf{a}^{(2)} &= (0, 0, \cdots\cdots\cdots, a_{n_1+1}, \cdots, a_{n_1+n_2}, 0, 0, \cdots \quad \cdots, 0) \\ &\vdots \\ \mathbf{a}^{(h)} &= (0, 0, \cdots \quad \cdots, 0, 0, a_{n_1+\cdots+n_{h-1}}, \cdots \\ &\qquad \cdots, a_r, 0, 0, \cdots, 0) \end{aligned} \tag{4.3.13}$$

are the left eigenvectors of P associated with eigenvalue 1, where n_i is the number of states in E_i for i = 1, 2, ···, h.

In general, the multiplicity of eigenvalue 1 of P is the number of ergodic classes of P. The corresponding right eigenvector $\mathbf{x}^{(i)}$ $(x_1^{(i)}, x_2^{(i)}, \cdots$ $\cdots, x_n^{(i)})$ of $\mathbf{a}^{(i)}$ is;

$$x_j^{(i)} = \begin{cases} 1 & \text{for } j \in E_i \\ 0 & \text{for } j \notin E_i, \quad \text{for } i = 1, 2, \cdots, h. \end{cases} \tag{4.3.14}$$

Suppose that an ergodic class, say E_1, is cyclic with period d. Then, by renumbering the states in E_1, P_1 has the following form.

$$P_1 = \begin{pmatrix} 0 & Q_1 & 0 & \cdots & 0 \\ 0 & 0 & Q_2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots\cdots & & Q_{d-1} \\ Q_d & 0 & \cdots\cdots & & 0 \end{pmatrix}. \tag{4.3.15}$$

Let $C_1, C_2, \cdots, C_d$ be the moving classes in E_1:

$$C_1 = \{1, \cdots, n_1'\},$$

$$C_2 = \{n_1'+1, \cdots, n_1'+n_2'\},$$

$$\vdots$$

$$C_d = \{n_1 - n_d' + 1, \cdots, n_1\}, \tag{4.3.16}$$

where n_i' is the number of states in C_i for i = 1, 2, ···, d.
From the definition of cyclic class, P_1^d has the form:

$$P_1^d = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & 0 \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & A_d \end{pmatrix}. \tag{4.3.17}$$

Namely, the Markov matrix P_1^d has d ergodic classes. Hence P_1^d has left eigenvectors $\mathbf{y}^{(1)}$, $\mathbf{y}^{(2)}$, ···, and $\mathbf{y}^{(d)}$ such that

$$\mathbf{y}^{(i)} \quad (0, 0, \cdots, y_1^{(i)}, y_2^{(i)}, \cdots, y_{n_i'}^{(i)}, 0, \cdots, 0)$$

$$\text{for } i = 1, 2, \cdots, d. \tag{4.3.18}$$

Let ω be the d^{th} root of 1, i.e., $\omega = e^{2\pi i/d}$. Then we have the following d equalities:

$$\mathbf{z}^{(k)} P_1 = \omega^{-k} \mathbf{z}^{(k)} \quad \text{for } k \quad 0, 1, 2, \cdots, d-1, \tag{4.3.19}$$

where

$$\mathbf{z}^{(k)} \quad \sum_{j=1}^{d} \omega^{k(j-1)} \mathbf{y}^{(j)}. \tag{4.3.20}$$

Let $\mathbb{Z}_k$, k = 1, 2, ···, d, be the n-dimensional vectors as follows:

$$\mathbb{Z}_k \quad (\mathbf{z}^{(k-1)}, 0, 0, \cdots, 0). \tag{4.3.21}$$

Then we have

$$\mathbb{Z}_k P = \omega^{-k+1} P, \text{ for } i = 1, 2, \cdots, d. \tag{4.3.22}$$

Therefore, if P has an ergodic class with period d, then P has eigenvalues 1, ω, ···, ω^{d-1}, and the corresponding left eigenvectors are given by (4.3.20) and (4.3.21).

4.3.3 Investigation of Structural Dependence of Quantification Vectors on Markov Chains Through Graph Theoretical Notions

A Markov matrix(Markov chain) P can be interpreted as a directed graph D_p (V_p, L_p) in the following way, where V_p is the set of the vertices of D_p and L_p is the set of the directed lines of D_p(cf. Harary, Norman, & Cartwright[1965]

and Iwahori[1974]). The set V_P $\{v_1, v_2, \cdots, v_n\}$ is the set of the states of Markov chain P, and the set L_P is given by

$$v_i v_j \in L_P \quad \text{iff} \quad p_{ij} > 0 \quad \text{for } i, j \quad 1, 2, \cdots, n. \tag{4.3.23}$$

When $v_i v_j$ is in D_P(i.e., $p_{ij} > 0$), we also write $v_i L_P v_j$. The binary relation " $\rightarrow$ " in Section 4.3.1 can be rewritten as

$$\begin{aligned} &i \rightarrow j \text{ iff there exists a path from } v_i \text{ to } v_j \text{(in } D_P\text{), i.e., there exist} \\ &\text{vertices } v_i = v_{q_1}, v_{q_2}, \cdots, v_{q_s} = v_j \text{ such that } v_{q_1} L_P v_{q_2}, v_{q_2} L_P v_{q_3}, \\ &\cdots, v_{q_{s-1}} L_P v_{q_s}. \end{aligned} \tag{4.3.24}$$

In this case, we also write it as $v_i L_P^* v_j$.

The graph theoretical notions below give the way of understanding the classification of states introduced in Section 4.3.1(for details, refer to Harary, Norman, & Cartwright[1965]). A *digraph*(directed graph) D = (V, L) is called *symmetric* iff $v_i v_j \in L$ means $v_i v_j \in L$. D is said to be reflexive iff $v_i v_j \in L$ for any $v_i \in V$. A digraph D is called *transitive* iff $v_i v_j \in L$ and $v_j v_k \in L$ means $v_i v_k \in L$ for all v_1, v_j, and v_k in V. D is called *strongly connected*, or (simply) *strong*, iff for every two vertices v_i and v_j of D, $v_i L^* v_j$ and $v_j L^* v_i$. Let V' be a subset of V. We call the maximal subgraph of D containing V' the *restriction* of D to V' and denote it by $D|_{V'}$, i.e.,

$$\begin{aligned} &D|_{V'} \quad (V', L'), \\ &L' = \{v_i v_j \mid v_i v_j \in L, v_i \in V', v_j \in V'\}. \end{aligned} \tag{4.3.25}$$

A *strong component* S of a digraph D is a subset of V such that $D|_S$ is strong and $D|_{S'}$ is not strong for any subset S'($\supsetneq$ S) of V.

The above notion of strong component partitions the vertex set V_P of D_P into the classes of strong components S_1, S_2, $\cdots$, and S_q. This yields the *condensation graph* $\overline{D}_P$ of D_P. The graph $\overline{D}_P$ is a digraph whose vertex set $\overline{V}_P$ is equal to $\{S_1, S_2, \cdots, S_q\}$ and directed lines are determined by the following rule: There is a line from vertex S_i to vertex S_j in $\overline{D}_P$ iff there exists at least one line from a vertex in S_i to a vertex in S_j in the graph D_P.

On the above condensation graph $\overline{D}_P$, we introduce the next classification of vertices due to Harary, Norman, & Cartwright(1965). An *isolate* is a vertex whose outdegree and indegree are both 0. A *transmitter* is a vertex whose outdegree is positive and whose indegree is 0. A *receiver* is a vertex whose outdegree is 0 and whose indegree is positive. A *carrier* is a vertex whose

outdegree and indegree are both 1. Any other vertex is called an *ordinary* vertex. In the above definition, the indegree(outdegree) of a vertex is the number of the lines going into(from) the vertex. It is easy to verify the following lemma.

LEMMA 4.3.1 : An ergodic class of P is a receiver or an isolate of $\overline{D}_P$ and *vice versa*. Hence any other kind of vertices of $\overline{D}_P$ corresponds to a set of transient states of P.

For example, let P be as follows:

$$P = \begin{pmatrix}
* & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & * & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & * & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & * & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0
\end{pmatrix}, \tag{4.3.26}$$

where *'s represent non-zero entities. Then the corresponding digraph is given by Fig. 4.3.1. The digraph has seven strong components S_1 S_7. The condensation graph $\overline{D}_P$ is shown by Fig. 4.3.2. S_1 is an isolate, S_2 and S_5 are transmitters, S_3 is a carrier, S_4 is an ordinary vertex, and S_6 and S_7 are receivers of $\overline{D}_P$. S_6 is an ergodic class of P consisting of single absorbing state, and S_7 is a cyclic ergodic class of P with period three.

Let $\overleftarrow{V}_P(v_i)$, $\overrightarrow{V}_P(v_i)$, $\overleftarrow{V}_P^*(v_i)$, $\overrightarrow{V}_P^*(v_i)$ be the sets of vertices defined by

$$\begin{aligned}
&\overleftarrow{V}_P(v_i) \triangleq \{v_j \mid v_j L_P v_i\}, \\
&\overrightarrow{V}_P(v_i) \triangleq \{v_j \mid v_i L_P v_j\}, \\
&\overleftarrow{V}_P^*(v_i) \triangleq \{v_j \mid v_j L_P^* v_i\}, \\
&\overrightarrow{V}_P^*(v_i) \triangleq \{v_j \mid v_i L_P^* v_j\}, \quad \text{for } i = 1, 2, \cdots, n.
\end{aligned} \tag{4.3.27}$$

Of course, it follows that

$$\overrightarrow{V}_P(v_i) \subset \overrightarrow{V}_P^*(v_i),$$

$$\overleftarrow{V}_P(v_i) \subset \overleftarrow{V}_P^*(v_i), \quad \text{for } i = 1, 2, \cdots, n. \tag{4.3.28}$$

Using the above notations, we have the following statements:

LEMMA 4.3.2 : The ergodic classes $E_1, E_2, \cdots, E_r$ of P are characterized by

$$E_i = \overrightarrow{V}_P^*(v_i) \quad \text{for any } v_i \in E_i \text{ and } i = 1, 2, \cdots, r. \tag{4.3.29}$$

The cyclic parts $C_1, C_2, \cdots, C_d$ in an ergodic class of P are characterized by

$$C_{k+1} = \bigcup_{v \in C_k} \overrightarrow{V}_P(v),$$

$$C_k = \bigcup_{v \in C_{k+1}} \overleftarrow{V}_P(v), \quad \text{for } k \quad 1, 2, \cdots, d \text{(we set } C_{d+1} = C_1). \tag{4.3.30}$$

PROOF : The lemma can be easily verfied from lemma 4.3.1, the definitions of ergodic class and cyclic part, and also (4.3.27). □

In the above, we characterized the classification of states of P in graph theoretical notions. In the sequel, we consider the relationship between the classification of states of U and that of P by the aid of the above graph theoretical notions.

As stated in Section 4.2.1, the Markov matrix U is difined by

$$U = (u_{ij}), \quad u_{ij} = \sum_k p_{ik}p_{jk}\pi_j \Big(\sum_j \pi_j p_{jk}\Big)^{-1} \tag{4.3.31}$$

Let $D_U = (V_U, L_U)$ be the corresponding digraph of U. Then, of course, we have

$$V_U = V_P = \{v_1, v_2, \cdots, v_n\}. \tag{4.3.32}$$

Let $V_{\boldsymbol{\pi}} (\subset V_U)$ be as follows:

$$V_{\boldsymbol{\pi}} = \{v_j \mid \pi_j > 0\}. \tag{4.3.33}$$

Then, from (4.3.31), we obtain

$$v_i v_j \in L_U \quad \text{iff} \quad \overrightarrow{V}_P(v_i) \cap \overrightarrow{V}_P(v_j) \neq \phi \text{ and } v_j \in V_{\boldsymbol{\pi}}. \tag{4.3.34}$$

That is, D_U is dependent not only on transition probabilities but also on initial distribution vector $\boldsymbol{\pi}$.

The next lemma states the dependence of D_U on $\boldsymbol{\pi}$.

LEMMA 4.3.3 : The vertices of $V_U - V_{\boldsymbol{\pi}}$ are transmitters of D_U. The digraph $D_U|_{V_{\boldsymbol{\pi}}}$, the restriction of D_U to $V_{\boldsymbol{\pi}}$, is symmetric and reflexive.

PROOF : From (4.3.34), we have

$$\text{If } \pi_j = 0 (\text{i.e., } v_j \notin V_{\pi}), \text{ then } v_i v_j \notin L_U \text{ for all } i = 1, 2, \cdots, n. \tag{4.3.35}$$

Because of U being a Markov matrix as defined in Section 4.2.1, there exists at least one k such that $u_{jk} > 0$, i.e., $v_j v_k \in L_U$. Hence, if $\pi_j = 0$, then v_j is a transmitter of D_U. If v_i and v_j are contained in V_{π}, then (4.3.34) yields

$$v_i v_j \in L_U \quad \text{iff} \quad \vec{V}_P(v_i) \cap \vec{V}_P(v_j) \neq \phi. \tag{4.3.36}$$

Hence we have

$$v_i v_j \in L_U \quad \text{iff} \quad v_j v_i \in L_U \text{ for all } i, j = 1, 2, \cdots, n. \tag{4.3.37}$$

$$v_i v_i \in L_U \quad \text{for all } i \quad 1, 2, \cdots, n. \tag{4.3.38}$$

Therefore, $D_U|_{V_{\pi}}$ is symmetric and reflexive. □

However, $D_U|_{V_{\pi}}$ is not necessarily transitive, for $\vec{V}_P(v_i) \cap \vec{V}_P(v_j) \neq \phi$ and $\vec{V}_P(v_j) \cap \vec{V}_P(v_k) \neq \phi$ do not assure $\vec{V}_P(v_i) \cap \vec{V}_P(v_k) \neq \phi$.

From lemmas 4.3.2 and 4.3.3, we finally obtain the following theorem which states the relationship between the classification of states of U and that of P.

THEOREM 4.3.1 : The following statements hold for an arbitrary Markov matrix P and probability vector π.

1) The transient class of U is given by $\{j \mid \pi_j = 0\}$.
2) U has no cyclic classes.
3) If the transient class of P is empty, then each ergodic class of U is contained in an ergodic class of P. Moreover, if an ergodic class of U is contained in a cyclic ergodic class of P, then it is contained in a cyclic part of the ergodic class of P.

PROOF : It is obvious that the transmitters of D_U are transient states of U. The digraph $D_U|_{V_{\pi}}$ is symmetric(cf. lemma 4.3.3), hence there exist no transient states in V_{π}. Hence the transient states of U are given by $V_U \quad V_{\pi}$, i.e. by $\{j \mid \pi_j = 0\}$.

This also means that the ergodic classes(and particularly the cyclic classes) of U are contained in V_{π}. From lemma 4.3.3, $D_U|_{V_{\pi}}$ is also reflexive, hence every ergodic class of U has period 1, i.e., U has no cyclic classes.

Let E_1, E_2, ···, and E_r be the ergodic classes of P. Then, from (4.3.36), lemma 4.3.2, and (4.3.28), we have

$$vv' \notin L_U \quad \text{for any } v \in E_i, \text{ any } v' \in E_j, \text{ any } j \text{ and } j' \text{ with } j \neq j'. \tag{4.3.39}$$

If the transient class of P is empty, i.e., $V_P = \bigcup_{k=1}^{r} E_k \ (= V_U)$, then from the above relation, we obtain

$$\vec{V}_U^{\,*}(v) \subset E_i \quad \text{for any } v \in E_i \text{ and any } i = 1, 2, \cdots, n. \tag{4.3.40}$$

From lemma 4.3.2, it follows that any ergodic class of U is contained in an ergodic class of P. Let $C_1, C_2, \cdots, C_d$ be the cyclic parts in an ergodic class of P. Then, from lemma 4.3.2,

$$\vec{V}_P(v_j) \subset C_{j+1} \quad \text{for } j = 1, 2, \cdots, r \text{ and any } v_j \in C_j. \tag{4.3.41}$$

It follows that

$$\vec{V}_P(v_j) \cap \vec{V}_P(v_{j'}) = \phi \quad \text{for any } v_j \in C_j, \text{ any } v_{j'} \in C_{j'}, \text{ and any } j \text{ and } j' \text{ with } j \neq j'. \tag{4.3.42}$$

Hence, if v_j is contained in a cyclic part C_j of P, then we have

$$\vec{V}_U^{\,*}(v_j) \subset C_j \quad \text{for any } v_j \in C_j \text{ and any } j = 1, 2, \cdots, d. \tag{4.3.43}$$

Therefore, if an ergodic class of U is contained in a cyclic ergodic class E of P, then it is contained in a cyclic part of E. □

For example, let Markov matrix P be given as follows:

$$P = \begin{pmatrix} * & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & * & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 \end{pmatrix}. \tag{4.3.44}$$

Then the corresponding digraph is shown by Fig. 4.3.3. In the figure, v_1 and v_2 are transient states of P, and E_1, E_2, and E_3 are ergodic classes. If the initial distribution vector π is given as

$$\pi = (*, *, \cdots, *), \tag{4.3.45}$$

then the Markov matrix U has the digraph as shown by Fig. 4.3.4. It shows that U has six ergodic classes E_1' E_6'. Ergodic class E_1' is composed of v_1 and v_2(

transient states of P), v_5 and v_6(which are contained in the ergodic class E_1 of P), and v_{10}(which is contained in the ergodic class E_3 of P). If π is given as

$$\pi = (0, *, *, 0, *, 0, *, *, *, 0, *), \tag{4.3.46}$$

then the corresponding digraph D_U is given by Fig. 4.3.5. The states v_1, v_4, v_6, and v_{10} with the initial probability 0 are the transient states of U. If P has the digraph D_P shown by Fig. 4.3.6 and the initial distribution is

$$\pi_i > 0 \quad \text{for all } i, \tag{4.3.47}$$

then D_U is given by Fig. 4.3.7. The ergodic classes E_1', E_2', and E_3' of U are contained in the ergodic class E_1 of P. The ergodic class E_2 of P is cyclic with cyclic parts C_1, C_2, and C_3. The ergodic classes E_4' and E_5' are contained in C_1. E_6' and E_7' coincide with C_2 and C_3, respectively.

Briefly speaking, the above theorem says that the transient class R and the recurrent class T of Markov chain U are determined only by the initial distribution, and if both P and U have no transient classes, then there exists a close relationship between the classification of states of U and that of P.

In the following, we consider the Markov chain W. In Section 4.2.1, $W = (w_{ij})$ is given as

$$w_{ij} = \sum_k \pi_k p_{ki} p_{kj} \left(\sum_m \pi_m p_{mi} \right)^{-1} \quad \text{for } i, j = 1, 2, \cdots, n. \tag{4.3.48}$$

The corresponding digraph $D_W = (V_W, L_W)$(,where $V_W = V_P = V_U$) has the following properties.

LEMMA 4.2.4 : D_W is symmetric and reflexive.

PROOF : From (4.3.48), we have

$$v_i v_j \in L_W \quad \text{iff } \overleftarrow{V}(v_i) \cap \overleftarrow{V}(v_j) \cap V_\pi \neq \phi \text{ for all } i, j = 1, 2, \cdots, n. \tag{4.3.49}$$

Hence D_W is symmetric. The condition (4.2.13) introduced in Section 4.2.1 is equivalent to

$$\overleftarrow{V}(v_i) \cap V_\pi \neq \phi \quad \text{for all } i \quad 1, 2, \cdots, n. \tag{4.3.50}$$

That is, D_W is reflexive. □

The above lemma leads to the next theorem which states the relationship between the classification of states of Markov chain W and that of P.

THEOREM 4.3.2 : The following statements hold for an arbitrary Markov matrix P and an arbitrary probabiliry vector π.

1) W has no transient class and no cyclic classes.
2) If the transient class of P is empty, then each ergodic class of W is contained in a cyclic ergodic class of P. Moreover, if an ergodic class of W is contained in a cyclic ergodic class of P, then it is contained in a cyclic part of the ergodic class of P.

PROOF : From the fact that D_W is symmetric and reflexive(by lemma 4.3.4), W has no transient and no cyclic classes. Let E_1, E_2, $\cdots$, and E_r be the ergodic classes of P. Then, from (4.3.49), lemma 4.3.2, and (4.3.28), we obtain

$$vv' \notin L_W \quad \text{for any } v \in E_j, \text{ any } v' \in E_{j'}, \text{ any } j \text{ and } j' \text{ such that } j \neq j'. \tag{4.3.51}$$

If the transient class of P is empty, i.e., $V_P = \bigcup_{k=1}^{r} E_k \ (= V_W)$, then from (4.3.51), we have

$$\vec{V}_W^*(v) \subset E_i \quad \text{for all } v \in E_i \text{ and all } i = 1, 2, \cdots, n. \tag{4.3.52}$$

From lemma 4.3.2, it follows that any ergodic class of W is contained in an ergodic class of P. Let C_1, C_2, $\cdots$, and C_d be the cyclic parts in an ergodic class of P. Then, from lemma 4.3.2,

$$\overleftarrow{V}_P(v_j) \subset C_{j-1} \quad \text{for } j = 1, 2, \cdots, r \text{ and any } v_j \in C_j \text{ (we set } C_0 = C_d). \tag{4.3.53}$$

It follows that

$$v_j v_{j'} \notin L_W \quad \text{for any } v_j \in C_j, \text{ any } v_{j'} \in C_{j'}, \text{ any } j \text{ and } j' \text{ such that } j \neq j'. \tag{4.3.54}$$

Hence, if v_j is contained in a cyclic class c_j of P, then we have

$$V_W^*(v_j) \subset C_j \quad \text{for any } v_j \in C_j \text{ and any } j = 1, 2, \cdots, d. \tag{4.3.55}$$

Therefore, if an ergodic class of W is contained in a cyclic ergodic class E of P, then it is contained in a cyclic part of E. □

Note that statement 3 of theorem 4.3.1 and 2 of theorem 4.3.2 are the same, however, the ergodic classes of U and those of W are not necessarily the same. For example, if P is given as Fig. 4.3.6 and π is given as (4.3.47), then D_W is is given as Fig. 4.3.8(cf. Fig. 4.3.7).

As stated in Section 4.3.2, the eigenvectors of a Markov matrix(Markov

chain) associating with the eigenvalues of absolute value 1 give the classification of states of the Markov chain. Also, as stated in Section 4.2.2, the eigenvalues of the Markov matrices U and W are real and nonnegative. Hence, the eigenvectors of U(W) associated with the maximum eigenvalue, i.e. 1, give the classification of states of U(W). Also, as shown in Section 4.2.1, the quantification vectors $\boldsymbol{\rho}$ and $\boldsymbol{\nu}$ for a Markov chain P are given as the eigenvectors of U and W associated with the maximum eigenvalues, respectively.

In Section 4.3, the relationship between the structural properties of original data(respose probability matrix P) and the quantification vectors $\boldsymbol{\rho}$ and $\boldsymbol{\nu}$ was discussed. From theorems 4.3.1 and 4.3.2, we can say that if a Markov chain(Markov matrix) P has such a certain complex structure as composed of multiple ergodic classes(some of which may be cyclic), if P has no transient classes, and if the initial distribution vector $\boldsymbol{\pi}$ satisfies (4.3.47)(i.e., if U has no transient class), then the quantification vectors $\boldsymbol{\rho}$ and $\boldsymbol{\nu}$ reflect the structure of P. That is to say, our quantification method tightly reflects the structural properties of original data.

4.4 Application of the Quantification Method to the Aggregation Problem of Inter-industrial Structures

In this section, we apply the quantification method introduced in Section 4.2 to the analysis of the inter-industrial structure of Japanese economy. In Section 4.4.1, we explain about the inter-industrial relations table(input-output table) and the method to transfer it into a Markov matrix. In Section 4.4.2, we discuss the relationships among industrial sectors of Japanese economy based on the quantification method. Particularly, the aggregation of the industrial sectors is discussed.

4.4.1 Input-output Table and Its Markov Matrix Version

An input-output table describes the flow of goods and services between all the industrial sectors of a national economy over a stated period of time-say, a year(for details, refer toLeontief[1941] and Kaneko[1976]).

There are many kinds of input-output tables depending on the treatment of imports and also on the purpose of analyses. We confine ourselves to the analysis of the table as shown by Table 4.4.1. In the table, the symbol x_{ij} stands for the amount of the product(in value terms) of sector i absorbed as an input by sector j. Each sector is supposed to produce individual goods. The symbol

x_{0j} represents the amount of production cost at sector j such as wages, depreciation expense, profit, and tax. F_i stands for the amount to the product of sector i absorbed as private or governmental consumptions, equipment investments, inventory investments, or exports. In the table, the sectors 1, 2, ···, and m are sometimes called *endogenous* sectors and the other sectors, Final demand and Gross value added, are called *exogenous* sectors. The rightmost column, Total output, represents the domestic output of each sector. The column sum in column j(containing Gross value added x_{0j}) coincides with the corresponding Total output T_j. The row sum in row i(containing Final demand F_i) is equal to Gross demand which is equal to the sum of (domestic) Total output T_i and the amount of imports I_i, i.e.,

$$G_i \quad T_i + I_i, \tag{4.4.1}$$

where G_i is the Gross demand(Gross supply) of sector i given by

$$G_i \quad \sum_{j=1}^{m} x_{ij} + F_i \text{ for } i = 1, 2, \cdots, m. \tag{4.4.2}$$

We transform the above input-output table into a response table composed of a single item with m + 1 categories. To make the table represent a closed flow of goods, we add the amount of imports I_i to Gross value added x_{0i}. Namely, the response matrix $Z = (z_{ij})$ is given as follows:

$$\begin{aligned}
&z_{ij} = x_{ij} \quad \text{for } i, j = 1, 2, \cdots, m, \\
&z_{i,m+1} = F_i \quad \text{for } i \quad 1, 2, \cdots, m, \\
&z_{m+1,j} = x_{0j} + G_j \quad F_j \quad \text{for } j = 1, 2, \cdots, m, \\
&z_{m+1,m+1} = 0.
\end{aligned} \tag{4.4.3}$$

In the above response table, the row sum of row i and the column sum of column i coincides with each other and are equal to G_i, the Gross demand(Gross supply) of sector i.

Hence the response matrix Z represents the closed flow of goods. The corresponding Markov matrix P and the initial probability distribution vector π are given by (4.2.2), (4.2.3), and (4.2.4).

4.4.2 Analysis of Inter-industrial Relationships based on the Quantification Method

In this study, we made the input-output table based on the third report issued from the Econometric Committee in Economic Deliberation Council of Japan in 1970(prediction value for the Japanese economy in 1975). This table is composed of 20 endogenous sectors shown by Table 4.4.2. Based on the input-output table, we obtain the Markov matrix P and the initial distribution vector $\boldsymbol{\pi}$ as shown by Table 4.4.3 and Table 4.4.4, respectively.

The quantification vectors $\boldsymbol{\nu}$ and $\boldsymbol{\rho}$, i.e., the right eigenvectors of U and W associated with the maximum eigenvalue $(\kappa_{\rho\nu})_{max}$(cf. (4.2.20)), are shown in Table 4.4.5. The eigenvalues $\kappa_{\rho\nu}$'s of equation (4.2.20) are shown by Table 4.4.6.

In this case, $(\kappa_{\rho\nu})_{max}$ 0.46425. (In Table 4.4.6, 1.00 corresponds to the meaningless case: ρ_i ν_i = 1 for all i = 1, 2, ···, 21.) It is easily shown that the response probability matrix P given by Table 4.4.3 constitutes a regular Markov chain, hence, the discussions in Section 4.3 have no relevance to this example. Also, by inspecting the matrix P, it will be immediately understood that the exogenous sector 21 has a special role, for most of the components of the 21st column in P are composed of extremely high probabilities. It It follows closely that, in the quantification vector, the 21st component is distinguished from the other components and most of the other components have nearly the same value.

To eliminate the effect of the exogenous sector(the 21st sector), we confine our analysis to the interrelationships among the endogenous sectors. In the Markov matrix P given by Table 4.4.3, we drop the 21st row and 21st column and then reconstruct it in such a way that each row sum equals 1. The initial probability vector $\boldsymbol{\pi}$ is also modified in the similar manner. The new transition probability matrix P is given by Table 4.4.7, and the new $\boldsymbol{\pi}$ is given as Table 4.4.8.

The new quantification vectors $\boldsymbol{\nu}$ and $\boldsymbol{\rho}$ are given by Table 4.4.9. The eigenvalues are shown by Table 4.4.10. In this case, the maximal correlation $(\kappa_{\rho\nu})_{max}$ = 0.52337 is considerably high and, hence, the associated eigenvectors $\boldsymbol{\nu}$ and $\boldsymbol{\rho}$ are supposed to reflect the inter-industrial structure of Japanese economy. To depict the above argument, we represent each sector i by the point (ν_i, ρ_i) in two-dimensional plane(cf. Katai, Imanaga, & Iwai[1975]). Fig. 4.4.1 shows the above representation. In the figure, the axis ν represents ν_i, the quantified value of each sector i in terms of its producing characteristics, and axis ρ represents ρ_i, the quantified value of sector i in terms of its consuming charcteristics. From the figure, we can find that the industrial sectors of Japan are broadly aggregated into three aggregations G_1 {1, 3, 4, 5, 6}, G_2

{7, 8, 9, 10, 11}, and G_3 {2, 12, 13, 14, 15, 16, 17, 18, 19, 20}. Moreover, it seems that aggregation G_1 is composed of two subaggregations G_{11} = {1, 3} and G_{12} = {4, 5, 6}, and aggregation G_3 is composed of two subaggregations G_{31} = {12} and G_{32} = {2, 13, 14, 15, 16, 17, 18, 19, 20}.

The group G_1 is the aggregation of the non-metallic sectors. G_2 is the aggregation of metallic sectors. The aggregation G_3 is composed of service or energy sectors; sector 12(Miscellaneous manufacturing) and sector 13(Construction) are not service or energy sectors, however the supply pattern of these two sectors are not partial to particular sectors, i.e., their products are used in most of the sectors, and hence, they are similar to service or energy sectors.

In aggregation G_1, subaggregation G_{11} is composed of the industries which are related to food and whose products are not intermidiate products absorbed by other industrial sectors, whereas the products of the industries in G_{12} are mainly intermidiate products and are absorbed by other insustrial sectors.

In aggregation G_3, the subaggregation G_{31}(Miscellaneous manufacturing) is distinguished from the other sectors in G_3 in the sense that it receives the goods which are produced in the sectors of G_1 and whose amounts are quite large. In the above analysis, it is easily seen that the quantification vector $\boldsymbol{\rho}$ is closely related to the aggregations and the subaggregations.

We next rearrange the order of 20 sectors according to the values of ρ_i's. Table 4.4.11 shows the transition probability matrix derived from the above rearrangement. In the table, each numeral 1 or 0 means that the corresponding transition probability is higher or lower than the average value, respectively. Also, in the table, the solid and the broken lines represent the division of the aggregations G_1, G_2, G_3, and the subaggregations G_{11}, G_{12}, G_{31}, and G_{32}. It will immediately be realized that the 0-1 matrix is composed of subregions(given by the division lines) where most of the numerals are 0 or subregions composed of many 1's. Hence the quantification vector $\boldsymbol{\rho}$ is quite in accordance with the actual flow in the Japanese economy.

In Section 4.4, the aggregation problem of Japanese inter-industrial structure was discussed through the quantification method introduced in Section 4.2. The analysis was based on the input-output table in 1975. We have clarified that the Japanese industries can be decomposed to three aggregations, i.e., metallic group, non-metallic group, and service or energy group. The anaysis was concerned only with the inter-industrial structure in 1975. If the analysis will be carried out for many years, then the trend of Japanese economical activity will be clarified, which will be beneficial for the developmental planning of Japanese economy.

4.5 Conclusions

In this chapter, we discussed the extension of the quantification method III to the case of probabilistic data. The problem of finding the quantification vectors ρ and ν in the extended method is reduced to an eigenvalue problem of certain Markov matrices as shown in Section 4.2. Also, in Section 4.2, it was shown that the correlation coefficient between the quantified values of sample types and those of category types has some desirable properties as a measure of dependence between the sample types and the category types. The analysis in Section 4.3 by graph theoretical methods has clarified that the quantification vectors ρ and ν introduced in Section 4.2 tightly reflect the qualitative properties of the original data. In Section 4.4, the extended method was applied to the aggregation problem of Japanese industrial sectors based on the input-output table in 1975. It was shown that Japanese industries can be aggregated into three groups, i.e. metallic group, non-metallic group, and energy or service group.

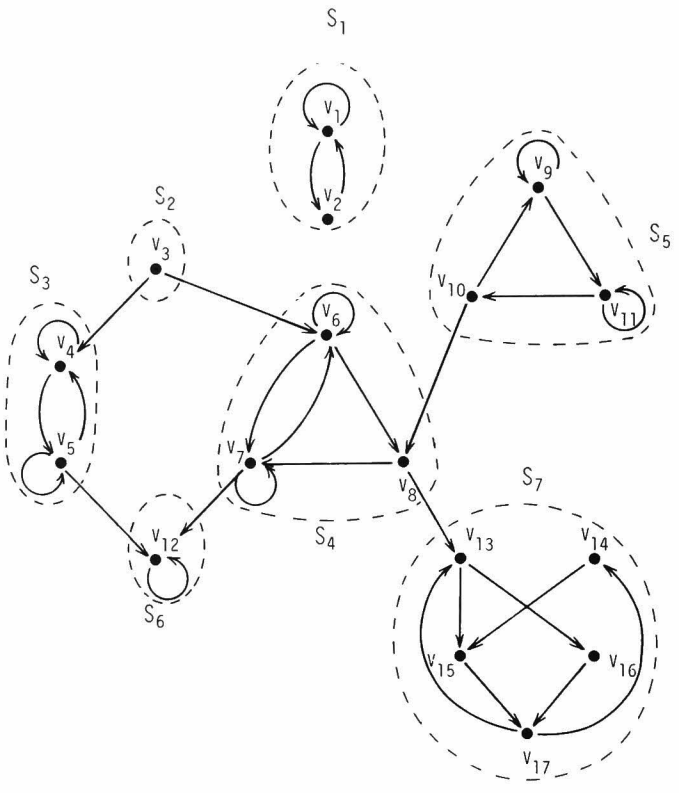


Fig. 4.3.1. Representation of the Markov chain prescribed by (4.3.26) as a digraph.

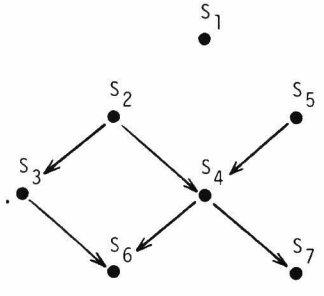


Fig. 4.3.2. Condensation graph of the digraph in Fig. 4.3.1.

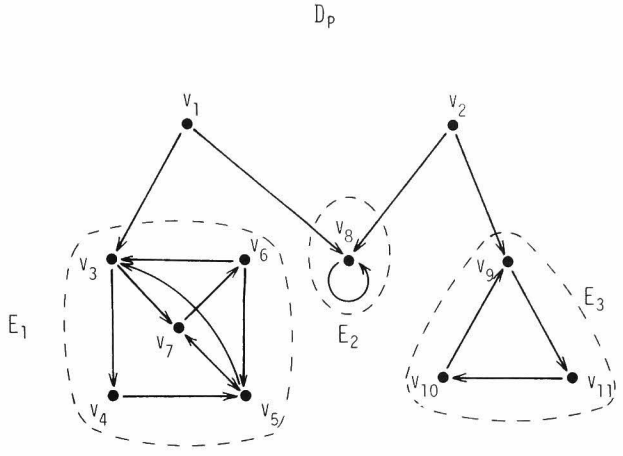


Fig. 4.3.3. The corresponding digraph D_P of the Markov matrix given by (4.3.33).

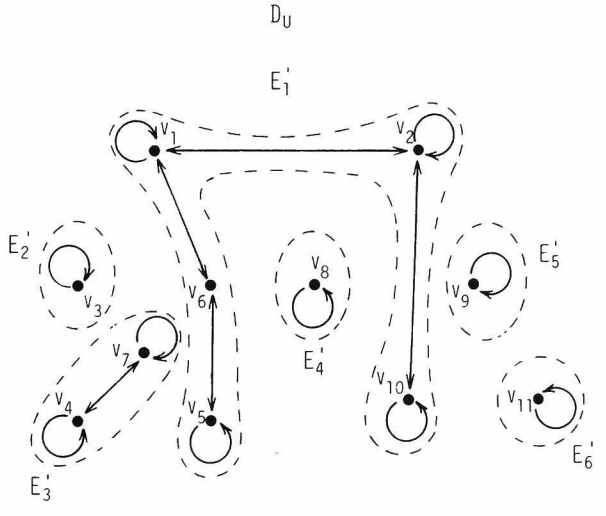


Fig. 4.3.4. The corresponding digraph D_U of the Markov matrix U associated with the Markov matrix P and the initian distribution vector π given by (4.3.45).

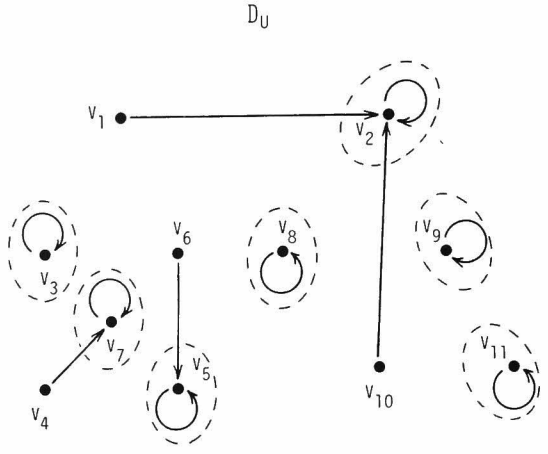


Fig. 4.3.5. The corresponding digraph of the matrix U associated with the initial distribution given by (4.3.46).

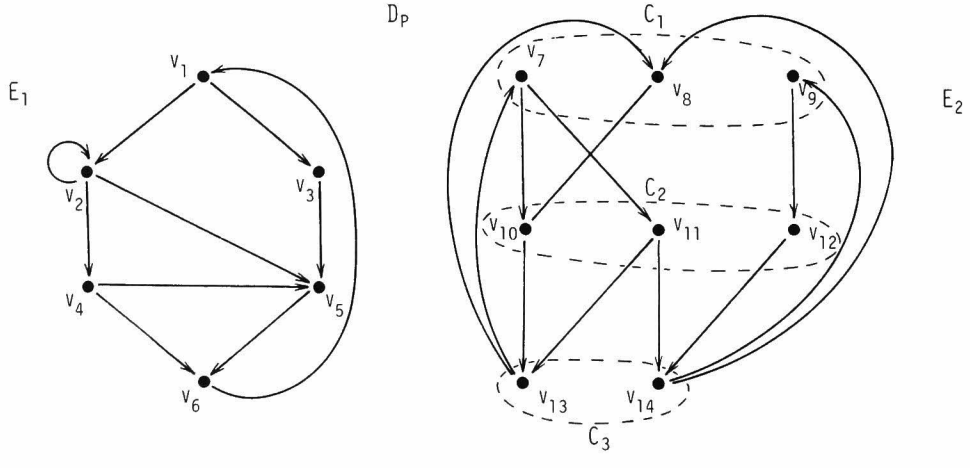


Fig. 4.3.6. A digraph representing a Markov chain composed of two ergodic classes one of which is consisted of three cyclic parts.

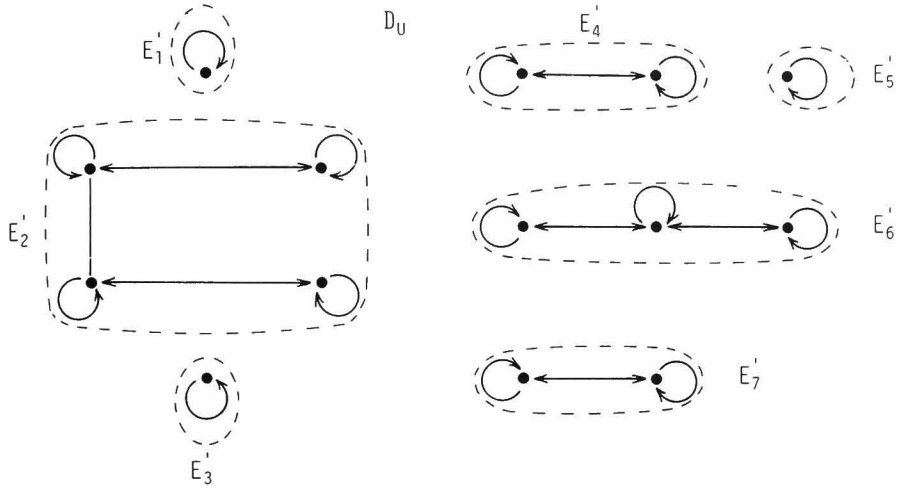


Fig. 4.3.7. The corresponding digraph of the Markov matrix U associated with the transition probabilities shown in Fig. 4.3.6 and the initial distribution vector satisfying (4.3.47).

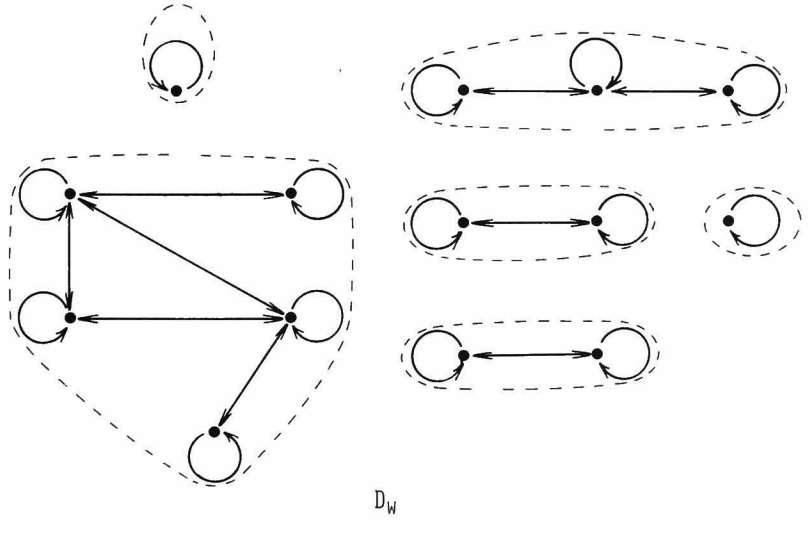


Fig. 4.3.8. The corresponding digraph of the Markov matrix W associated with the same Markov chain as in Fig. 4.3.7.

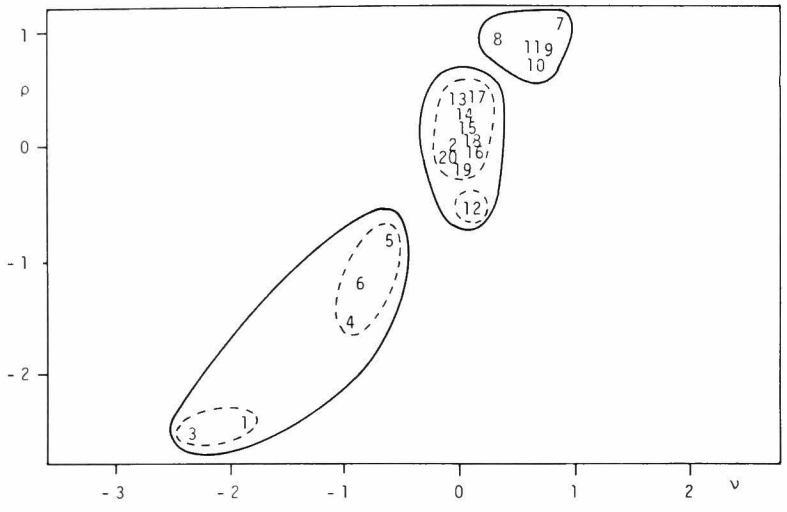


Fig. 4.4.1. Aggregation of the Japanese inter-industrial structure in 1975 by the use of quantifications ν and ρ for 20 endogenous sectors.

From \ Into	(industrial) sectors 1	2	.....	m	Final demand	Total output
(industrial) sectors 1	x_{11}	x_{12}	.....	x_{1m}	F_1	T_1
2	x_{21}	x_{22}	.....	x_{2m}	F_2	T_2
.	.				.	.
.	.				.	.
.	.				.	.
m	x_{m1}	x_{m2}	.....	x_{mm}	F_m	T_m
Gross value added	x_{01}	x_{02}	.....	x_{0m}	0	T_0

Table 4.4.1. An input-output table.

1: Agriculture, forestry, and fishery
2: Mining, petroleum, and natural gas
3: Processed foods
4: Textiles
5: Paper and pulp
6: Chemicals
7: Primary metals
8: Metal products
9: Machinery(except electrical machinery)
10: Electrical machinery
11: Transportation equipment
12: Miscelleneous manufacturing
13: Construction
14: Electricity, gas, and water supply
15: Trade
16: Real estate and rentals
17: Transportation and communications
18: Banking and insurance
19: Public and other services
20: Unallocated

Table 4.4.2. Contents of the 20 endogenous sectors.

$\times 10^4$

sectors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	653	4	3959	682	189	421	6	9	0	9	0	1429	129	0	0	0	0	0	31	78	2394
2	3	10	32	3	43	257	2679	4	5	30	1	5504	713	395	0	0	8	1	6	16	281
3	753	0	1468	11	2	85	0	0	0	0	0	126	0	0	17	0	0	0	1	58	7473
4	56	1	6	2600	16	16	22	28	78	119	47	2047	186	3	127	0	31	24	31	34	4520
5	11	4	149	21	4254	805	479	34	56	197	62	2795	142	6	481	0	27	32	110	23	302
6	212	11	321	700	97	3761	211	102	143	363	202	1597	151	8	12	0	9	0	526	76	1489
7	0	1	0	0	0	5	5212	974	742	823	610	209	522	6	1	0	0	0	1	34	850
8	32	6	190	2	18	119	106	939	849	321	629	544	3994	17	268	0	29	0	88	430	1409
9	10	4	1	9	5	13	102	29	2094	189	1325	59	315	27	6	0	4	0	10	6	5783
10	5	3	1	3	0	13	99	26	316	2532	429	25	728	208	0	0	20	0	27	13	5541
11	45	1	6	3	0	0	39	1	149	100	2719	22	307	0	535	0	669	0	23	12	5361
12	38	12	124	33	203	352	453	81	403	538	535	1487	1512	185	579	0	439	129	587	184	2116
13	30	9	26	6	4	32	41	41	36	34	42	72	17	119	204	122	48	46	188	0	8873
14	47	78	216	123	468	794	1356	188	411	462	493	775	172	134	535	6	265	125	988	89	2266
15	81	15	406	140	112	341	569	266	439	527	623	998	1130	61	417	0	149	66	298	191	3158
16	0	1	5	9	4	12	27	5	65	48	21	87	46	18	464	2	67	112	242	0	8754
17	67	13	218	56	133	264	626	194	417	331	368	795	1310	114	472	0	315	130	290	80	3799
18	61	16	120	59	57	115	232	143	286	327	368	527	435	23	2002	27	118	1978	344	362	2390
19	5	2	52	9	5	48	60	4	107	87	57	122	73	7	236	1	34	48	314	0	8719
20	148	145	601	1241	139	327	1537	0	942	0	908	495	0	266	0	32	1230	701	731	547	0
21	524	523	353	108	128	699	460	271	630	427	622	1214	347	199	1448	96	640	329	870	99	0

Table 4.4.3. Markov matrix version of the input-output table consisted of 20 endogenous sectors and one exogenous sector for the Japanese economy in 1975.

sectors	Π_i
1	0.02186
2	0.01565
3	0.03129
4	0.01488
5	0.01472
6	0.04967
7	0.07066
8	0.02118
9	0.04849
10	0.04458
11	0.06200
12	0.09007
13	0.05726
14	0.01107
15	0.06150
16	0.00355
17	0.03063
18	0.01590
19	0.04102
20	0.00927
21	0.28474

Table 4.4.4. The initial distribution vector associated with the Markov matrix shown in Table 4.4.3.

sectors	ν	$\wp$
1	-0.01299	0.76230
2	0.73942	1.39818
3	-1.00597	0.17988
4	-0.36541	0.37298
5	0.91927	1.26276
6	0.42305	0.96729
7	0.47273	0.90834
8	0.08554	0.83606
9	-0.80500	0.21753
10	-0.75936	0.08386
11	-0.72770	-0.14876
12	0.11882	0.84891
13	-1.24632	0.17968
14	0.20376	0.56484
15	-0.12859	0.83946
16	-1.21516	0.62177
17	-0.23245	0.68129
18	0.17078	0.91136
19	-1.21986	0.82725
20	0.58284	0.55742
21	0.67971	-1.49124

Table 4.4.5. The quantified values for the sectors.

	$\kappa_{\rho\nu}$
1	1.00000
2	0.46425
3	0.30776
4	0.19351
5	0.16315
6	0.11853
7	0.10048
8	0.06923
9	0.06676
10	0.05154
11	0.04319
12	0.03347
13	0.02643
14	0.00971
15	0.00353
16	0.00246
17	0.00101
18	0.00047
19	0.00019
20	0.00001
21	0.00000

Table 4.4.6. The eigenvalues of equation (4.2.20).

x 10^4

sectors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	858	6	5205	896	249	554	8	12	0	13	0	1878	169	0	0	0	0	0	41	103
2	3	10	32	3	44	264	2757	4	6	31	2	5664	734	406	0	0	8	1	6	17
3	2983	0	5812	46	10	339	0	0	0	0	0	501	0	0	69	0	0	0	5	229
4	103	3	12	4744	29	29	40	51	142	218	86	3737	339	7	232	0	56	44	57	62
5	11	4	154	21	4386	831	494	35	58	203	64	2882	147	6	496	0	28	33	114	24
6	249	13	377	822	114	4419	246	120	168	427	237	1876	178	9	14	0	11	0	619	89
7	0	1	0	0	0	6	5697	1065	811	900	667	229	571	6	1	0	0	0	1	37
8	37	7	221	3	21	138	123	1094	988	374	733	634	4650	20	312	0	34	0	103	501
9	25	10	4	21	13	31	242	70	4966	449	3142	140	747	66	14	0	10	2	25	14
10	12	8	2	7	0	31	223	60	710	5679	963	57	1634	468	0	0	46	0	62	31
11	97	2	14	7	0	3	84	3	321	217	4862	48	663	1	1153	0	1443	0	50	27
12	49	15	157	42	257	447	574	103	511	682	679	1887	1918	234	734	0	556	164	745	234
13	269	88	238	54	44	289	369	366	324	309	375	641	154	1060	1819	1084	429	410	1670	1
14	60	101	279	159	606	1027	1754	243	532	597	638	1002	222	173	692	7	343	162	1278	115
15	119	23	593	205	164	493	832	390	642	771	910	1459	1652	89	609	0	219	97	437	279
16	0	15	45	75	32	103	216	47	525	392	173	704	374	147	3733	16	543	901	1948	0
17	109	21	353	90	215	425	1010	313	672	534	594	1282	2112	184	762	0	508	210	468	129
18	81	21	157	77	75	152	305	188	376	430	484	693	572	30	2631	36	156	2600	452	475
19	41	22	410	70	40	382	472	32	840	685	451	954	573	61	1849	8	267	379	2452	2
20	148	145	601	1241	139	327	1537	0	942	0	908	495	0	266	0	32	1230	701	731	547

Table 4.4.7. The Markov matrix representing the internal flow of goods among 20 endogenous sectors.

sectors	Π_i
1	0.03862
2	0.03533
3	0.01836
4	0.01894
5	0.03316
6	0.09818
7	0.15017
8	0.04227
9	0.04749
10	0.04617
11	0.06679
12	0.16493
13	0.01498
14	0.01988
15	0.09774
16	0.00103
17	0.04412
18	0.02810
19	0.01220
20	0.02153

Table 4.4.8. The initial distibution vector associated with the Markov matrix shown by Table 4.4.7.

sectors	ν	ρ
1	-1.89573	-2.48352
2	-0.07494	-0.15294
3	-2.32133	-2.57294
4	-0.96449	-1.59896
5	-0.62669	-0.88426
6	-0.88285	-1.24997
7	0.85189	1.01172
8	0.30719	0.87078
9	0.71821	0.80878
10	0.58644	0.64834
11	0.58341	0.80759
12	0.03520	-0.60806
13	-0.05654	0.35574
14	-0.02705	0.21230
15	-0.02607	0.15286
16	0.06608	-0.10214
17	0.07819	0.37131
18	0.05308	0.01329
19	-0.04494	-0.27638
20	-0.14819	-0.16791

Table 4.4.9. The quantified values for the endogenous sectors based on the structure shown in Tables 4.4.7 and 4.4.8.

	$\kappa_{\rho\nu}$
1	1.00000
2	0.52237
3	0.36188
4	0.33744
5	0.24403
6	0.21235
7	0.19900
8	0.15377
9	0.12693
10	0.12208
11	0.10573
12	0.06604
13	0.03412
14	0.02275
15	0.01765
16	0.00312
17	0.00077
18	0.00058
19	0.00004
20	0.00002

Table 4.4.10. The eigenvalues of equation (4.2.20) for the case conditioned by the values in Tables 4.4.7 and 4.4.8.

		G_2					G_3								G_{31}	G_1				
							G_{32}									G_{12}			G_{11}	
		7	8	9	11	10	17	13	14	15	18	16	2	20	19	12	5	6	4	1	3
G_2	7	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
	8	0	1	1	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
	9	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G_3	17	1	0	1	1	1	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0
	13	0	0	0	0	0	0	0	1	1	0	1	0	0	1	1	0	0	0	0	0
	14	1	0	1	1	1	0	0	0	1	0	0	0	0	1	1	1	1	0	0	0
	15	1	0	1	1	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1
G_{32}	18	0	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0
	16	0	0	1	0	0	1	0	0	1	1	0	0	0	1	1	0	0	0	0	0
	2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
	20	1	0	1	1	0	1	0	0	0	1	0	0	1	1	0	0	0	1	0	1
	19	0	0	1	0	1	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0
G_{31}	12	1	0	1	1	1	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0
G_1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
G_{12}	6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
G_{11}	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1

Table 4.4.11. Quantized 0-1 matrix representing the internal flow of goods, where the sectors are ordered according to the values given by quantification ρ.

CHAPTER 5 | AGGREGATION OF SOCIOMETRIC GROUP STRUCTURES THROUGH QUANTIFICATION AND GRAPH THEORETICAL METHODS, AND QUANTIFICATION AND CLASSIFICATION OF DEGREE OF SOCIAL BALANCE THROUGH STATISTICAL AND FINITE-STATE SYSTEMS THEORETICAL ANALYSES

5.1 Introduction

In this chapter, we deal with the aggregation problems of sociometric group structures in relation to the theory of social balance developed by Cartwright and Harary. As mentioned in Chapter 1, sociometric group structures can be represented by the data of the fourth type, where the response matrix $A = (a_{ij})$ corresponds to the so-called *sociogram* introduced by Moreno(1934). This is represented by the following quantities:

$$a_{ij} = \begin{cases} 1 & \text{when the } i^{th} \text{ member chooses(likes) the } j^{th} \text{ member,} \\ 0 & \text{when the } i^{th} \text{ member has no relationships with the } j^{th} \text{ member,} \\ -1 & \text{when the } i^{th} \text{ member rejects(dislikes) the } j^{th} \text{ member,} \end{cases}$$

$$\text{for } i, j = 1, 2, \cdots, n. \qquad (5.1.1)$$

In this type of group structure, the aggregation means to decompose groups into tight or highly cohesive subgroups. A *clique* is the most typical of such a subgroup. The cliques of a group are defined as the subgroups such that there are no negative relations inside of each subgroup and no positive relations joining different subgroups. In social psychology, these cliques are considered as the most fundamental constituents of social groups(cf. Festinger[1949], Luce[1950], and Harary[1959]).

Sociometry, introduced by Moreno, is a method to analyze the internal structure of social groups for the sake of improvement of interpersonal relationships, of the productivity of social groups, or of individual member's life. This is done by finding the above mentioned cliques, examining the interrelationships among cliques, and classifing the types of group members into such categories as isolates, stars, and fringers(for details, refer to Ross & Harary [1955], Moreno[1960], and Tanaka[1975]).

There is another type of study for examining social group structures called *balance theory* introduced by Cartwright and Harary(1956). The theory intends

to interpret the behaviors of social groups by the following general rule: an unbalanced group tends to balance itself by changing the interpersonal relations inside the group(e.g., refer to Feather[1965], Tallman[1967], and Carroll[1973]).

The Cartwright-Harary theory is based on the general theory on cognitive balance of an individual person introduced by Heider(1946). They employed the graph theoretical representation of sociometric group structures, as mentioned in Chapter 1, and the definition of social balance is given in terms of the signs of the cycles in the graphs. This will be briefly reviewed in the next section.

They have shown that a balanced group is composed of either one or two cliques and an unbalanced group is one which has no cliques or more than two cliques. Hence, the aggregation of sociometric group structures, from the balance theoretical point of view, is interpreted as the decomposition of groups into no more than two tight or highly cohesive subgroups. For a balanced group, its aggregation means decomposition into the one or two cliques which constitute the group. Also, for an unbalanced group, aggregation means decomposition into at most two tight subgroups which are nearly cliqual, i.e. the most cohesive.

This aggregation problem for an unbalanced group structure can be reduced to finding out the most proximate balanced state to the original state, by regarding the tight subgroups as cliques. In other words, this problem is equivalent to the balancing problem of unbalanced social groups with minimum effort, i.e., by the use of the minimum number of alterations of member-member relationships.

In general, a set of member-member relationships in an unbalanced group whose alteration yields a balanced state is called a *balancing set* of the group, and a balancing set with the minimum number of elements is called a *minimum balancing set*. Hence, the above aggregation problem for an unbalanced group structure can be reduced to the derivation of the minimum balancing sets of the group. Also, the number of elements in a minimum balancing set of an unbalanced group can be interpreted as a measure of degree of balance of the group(cf. Abelson & Rosenberg[1958] and Flament[1963]). Hence, the aggregation method for unbalanced group structures also gives the way for calculating the degree of balance.

In the next section, we give a brief review of the theory of social balance developed by Cartwright and Harary. In Section 5.3, we deal with the aggregation problems of balanced and unbalanced group structures. First, a modified method of Hayashi's quantification method for the data of the fourth type is

introduced, and it is shown that the method can specify or detect the cliques in balanced group structures. Second, an iterative method based on factor analysis techniques is incorporated into the quantification method for the aggregation of unbalanced group structures. Because of the notion of balance being, in nature, graph theoretical, some graph theoretical methods are introduced and proved to be quite effective for the aggregation of unbalanced groups with graph structures of a special kind. Also, in this section, some extensions of the aggregation problems for unbalanced group structures are discussed. As mentioned before, these aggregation methods for unbalanced group structures provide a way for the quantification of degree of balance for these group structures. In actual social groups, however, their sociometric structures are sometimes inconstant. Hence, in Section 5.4, the degree of balance is quantified from a statistical point of view, and also characterization of types of social balance is discussed in view of finite-state systems theory.

5.2 Various Notions Concerning the Balance of Social Groups

The fundamental constituents of various social group structures, such as political and economical group structures, are the *units*(personal units or various social units such as nations, companies etc.) and the *objects*(impersonal objects such as social affairs) of common interest by the units. In such group structures, each unit has its *sentiments*(likes, dislikes), *opinions*(approval, disapproval) or *attitudes*(friendliness,hostility), about the objects and other units.

When a group structure is viewed through the above *cognitive relationships*, the problem of the *cognitive balance* of the whole group becomes of primal importance. An essential and simplified grasp of the cognitive balance is obtained by the following theory of Heider(1946). He considered the situation composed of two persons(units) P and O and an impersonal object X. He categorized the relations among P, O and X as positive(+) or negative(-), where + means to like or to approve or to have friendly relations and means to dislike or to disapprove or to have hostile relations. Fig. 5.2.1 shows four possible balanced states of the above *2-unit 1-object group* asserted by Heider, where solid and broken lines correspond to positive and negative relations, respectively. In Fig. 5.2.1 (a) and (b), both units P and O have similar sentiments(or opinions or attitudes) about the object X and so the relation between P and O is positive (friendly). In Fig. 5.2.1 (c) and (d), the sentiments of the two units about X are opposite and the relation between P and O is negative(hostile). Therefore,

no motivation to change the situation occurs.

If the signs of all the relations are reversed, the four balanced situations above are changed to unbalanced situations, and a motivation to restore their balance arises. For example, in Fig. 5.2.1 (a), if all the signs are changed to negative, units P and O will have a negative(hostile) relation with each other in spite of the coincidence of P and O's negative sentiments about the object X. This means that the group is unbalanced and the following four types of motivation to restore its balance arise.

(1) P and O change the relation P-O between them from to + to attain the balanced situation (b).

(2) P changes his sentiment P-X from to + to attain (c).

(3) O changes his sentiment O-X from to + to attain (d).

(4) P and O change their relations P-O, P-X, and O-X from to + to attain (a).

These processes of restoring balance are called the *balancing processes*. Case (4) requires the sign change three times, while the cases (1), (2), and (3) require only one sign change. Therefore, case (4) will not readily occur as an actual balancing process. Among the above four cases, cases (1), (2) and (3) are called the *minimum balancing processes*.

Heider's notion of the cognitive balance can be formally defined as the *sign of circuit* P-O-X-P being positive, where the sign is the product of the signs of the relations contained in the circuit and a circuit is defined as a closed path.

An extension of Heider's theory was made by Cartwright and Harary(1956) to treat more realistic social problems. Their modeling of the social group structures employs a general *signed graph*(*s-graph*), and no restrictions on the number of the units or objects are set. In their models, the *lines* connecting the *vertices* of the signed graphs represent the cognitive relations + or similar to Heider's model. But they do not refer to the distinction between the units and the objects as in Heider's model, and both of them are regarded as the members(vertices) of the groups(s-graphs).

One of the most interesting points suggested by Cartwright and Harary using their models is that, in a balanced group, the set of the vertices(units and objects) is aggregated into two subgroups(one of which may be empty) in such a way that the relations between the vertices of the same subgroup have positive signs and the relations between the vertices of different subgroups have negative signs. The above statement is referred to as the *Structure Theorem*. From the theorem, a balanced state means a *tight bipolar* configulation of the group, i.e.,

the situation of the group governed by two distinct and exclusive ideologies or value judgements (cf.Kaplan[1967]). Strictly speaking, the balance of a group is defined as the state in which all the signs of the cycles are positive in the corresponding signed graph. As previously mentioned, the sign of a cycle is the product of the signs of the lines (relations) contained in the cycle. The notion of social balance defined above is sometimes referred to as *structural balance*.

We will show a simple example to demonstrate the notion of the structural balance. Fig. 5.2.2 shows a group of three units (nations *USA*, *USSR* and *China (C)*) and two objects (Chinese Nuclear Armament *(C.N.)* and Intervention of *USSR* to African Countries *(AF)*). Fig. 5.2.2(a) is an unbalanced state because the cycles *USA-USSR-C.N.-USA* and *USA-AF-C-USA* have negative signs. When *USA* changes her attitude towards *USSR* and *AF* from to + as shown in Fig. 5.2.2(b), the group is balanced. In this case, the set of vertices are aggregated into {*USA*, *USSR*, *AF*} and {*C*, *C.N.*}, that is, *USA* and *USSR* have friendly relations by the concession of *USA* to *USSR* on the item *AF*, and as a consequence, *China* is isolated. The above situation is a tight bipolar configuration. In Fig. 5.2.2(c), *China* has conceded to *USSR* concerning *AF* and *USSR* has reciprocated by approving *C.N.*, changing the *USSR-C* relation to positive. This system is balanced in another bipolar configuration, *USA* being isolated from the group of *USSR* and *China*. In this case, the relations *USSR-C.N.*, *USSR-C* and *AF-C* have to change their signs. On the other hand, in case (b), only the relations *USA-USSR* and *USA-AF* must be changed. The process yielding situation (b) is the minimum balancing process, but the process yielding (c) is not. Fig, 5.2.2(d) shows a balanced state attained by another minimum balancing process, so, the minimum balancing process is not necessarily unique.

The above balancing processes can be characterized by the sets of the lines whose signs are changed during the processes. For example, the balancing process yielding (c) can be represented by the set: {*USSR-C.N.*, *USSR-C*,*AF-C*}. These sets are called balancing sets, and particularly, the sets corresponding to the minimum balancing processes are called minimum balancing sets.

Harary(1961) applied the Cartwright-Harary theory to the Suez Canal crisis in the Middle East. Kammler(1974) demonstrated how actual balancing processes take place, by using the ancient political situations between Rome, Tarentum, Syracuse and Carthage, and also, by using the situations in the European countries in the 18th century. These groups were motivated toward a

restoration of balance, as other groups will in the way previously explained.

The characterization of the balancing processes in the various types of groups with cognitive relations is of considerable importance for more precise understanding and forecasting of the future situations. In particular, the minimum balancing processes, which yield the most proximate balanced state by minimum sign change, i.e., minimum effort, is of actual importance. Another kind of balancing sets with actual importance is introduced by Harary(1959) and is called *minimal balancing sets*, which are defined as the balancing sets whose proper subsets are no longer balancing sets.

There are many indices for measuring the degree of balance of social groups. Cartwright and Harary (1956) proposed as a first approximation of degree of balance the ratio of the number c^+ of positive circuits to the total number c of circuits. Abelson & Rosenberg(1958) and Flament(1963) considered as the degree of balance (more precisely, degree of unbalance) the number of lines (relations) in the minimum balancing set, i.e. the smallest number of sign changes necessary to restore balance. Harary(1959) called the above number *line index* and also introduced the notion of *point index* which is defined as the smallest number of vertices whose deletion results in balance. Norman and Roberts(1972) extended the measure c^+/c by incorporating the effect of the length of circuits, and some generalizations of their work were made by Hansen(1975). Among the above definitions, the measure by Abelson and Rosenberg has a tight connection with the aggregation of unbalanced group structures, for the measure is defined in terms of the minimum balancing sets. Namely, the calculation of the measure is reduced to search for the minimum balancing sets of an s-graph.

5.3 Aggregation of Sociometric Group Structures into Tight Subgroups through Quantification and Graph Theoretical Methods

In this section, we deal with the aggregation problem of balanced or unbalanced group structures. Namely, we consider the method for aggregating balanced groups into cliques and that for the derivation of the minimum balancing sets for unbalanced groups.

Section 5.3.1 deals with the aggregation of balanced group structures, where a quantification method, which is a modification of Hayashi's quantification method IV, is introduced to the members of groups. It is shown that the members with positive quantified values and those with negative quantified values constitute the cliques in balanced group structures. However, for

unbalanced group structures, the above method does not necessarily specify the tight subgroups, i.e., is not necessarily available for the derivation of the minimum balancing sets. Hence some modifications, discussed in Section 5.3.2, should be incorporated into the above method. Namely, instead of attaching metric values to the members of groups as in the quantification method, we consider a method for attaching nominal values to the members. Precisely, the sign +1 or -1 is attached to each member. The members with the same sign are regarded as constituting a tight subgroup. In other words, the aggregation of an unbalanced group is characterized in terms of sign vectors. By the use of the sign vectors, it is shown that the above aggregation problem is reduced to a quantification problem of the members which is related to the centroid factor analysis techniques. Furthermore, an iterative method for the above quantification is introduced, in which the quantification method for the balanced group structures has a role in determining the initial values. While the above method is widely available for the aggregation of unbalanced group structures, it is to be expected that some specific methods are more effective for specific group structures. In Section 5.3.3, it is shown that graph theoretical methods are quite effective for the aggregation of unbalanced groups with planar graph structures. Sections 5.3.4 and 5.3.5 deal with the aggregation problem of unbalanced group structures in slightly different frameworks. In Section 5.3.4, the case where the member-member relationships have their own directions is discussed. The above treatments of the aggregation problem are based on the assumption that the relationships in groups are homogeneous; that is, each relation is of the same relative importance. In actual social groups, however, the relations have to be analyzed not only with respect to their signs but also to their actual relative importance. For example, in the 3 - nation 2 - affair group in Fig. 5.2.2, the relation between *USA* and *AF* may be of more actual importance than that between *C* and *AF*. Also, in Section 5.3.4, we deal with the problem for the group structures with inhomogeneous relationships. The treatment needs no essential alteration from that of the homogeneous case. We also show how our methods are applied to the problem of *coalition formation*. In Section 5.3.5, the aggregation problem is considered in a more general framework, i.e., we consider the derivation of the minimal balancing sets of unbalanced groups. As mentioned in Section 5.2, the notion of minimal balancing sets is an extension of that of the minimum balancing sets and is of actual importance. We consider the aggregation of unbalanced groups into subgroups which are a bit more loosely cohesive compared to the subgroups prescrib-

ed by minimum balancing sets.

5.3.1 Aggregation of Balanced Group Structures into Cliques through a Quantification Method

In balanced groups, the sociometric structures are necessarily symmetric in the following sense.

$$a_{ij} \quad a_{ji} \quad \text{for all } i,j \quad 1, 2, \cdots, n. \tag{5.3.1}$$

For, if $a_{ij} = -a_{ji}$, then the cycle composed of the lines v_iv_j and v_jv_i has a negative sign.

In this symmetric case, the matrix $B = (b_{ij})$ used in the quantification method IV is given as (cf. Section 2.3.2),

$$\left.\begin{aligned} b_{ij} &= 2a_{ij} \quad \text{for } i \neq j \\ b_{ii} &= -\sum_{\substack{j=1 \\ j\neq i}}^{n} b_{ij} \quad \text{for } i \quad 1,2, \cdots, n \end{aligned}\right\}. \tag{5.3.2}$$

Namely, B 2A except for the diagonal elements. (The diagonal element a_{ii} of A is assumed to be 0 for i = 1, 2, $\cdots$,n).

The quantification vector φ for the data of the fourth type is given as the eigenvector of the matrix B associated with the maximum eigenvalue, and is expected to specify the clustering of sociometric structures, as briefly reviewed in Section 2.3.2. However, as shown in the sequel, the eigenvector of the matrix A itself associated with the maximum eigenvalue is more adapted to the aggregation of balanced group structures.

Let S be a balanced s-graph S. Then the structure theorem says that the set V of the vertices of S can be decomposed into two groups V_1 and V_2 (one of which may be empty) such that any line joining the vertices of the same group is positive and any line joining the vertices of different groups is negative. When V_1 or V_2 is empty, then the adjacency matrix A(S) is a positive matrix, i.e. a matrix whose entities are nonnegative and at least one or them is positive. When neither V_1 nor V_2 are empty, then renumbering the vertices in S yields A(S) as

$$A(S) = \begin{pmatrix} A_1 & A_3 \\ {}^t A_3 & A_2 \end{pmatrix}, \tag{5.3.3}$$

where A_1 and A_2 are positive matrices and A_3 is a negative matrix, i.e., every entity of A_3 is nonpositive and at least one is negative.

The optimum quantification vector $\boldsymbol{\rho}^*$ is given as the eigenvector of A(S) associating with the maximum eigenvalue. Let $\boldsymbol{\rho}'$ be

$$\boldsymbol{\rho}' \triangleq \begin{pmatrix} \boldsymbol{\rho}_1^* \\ -\boldsymbol{\rho}_2^* \end{pmatrix}, \tag{5.3.4}$$

where $\boldsymbol{\rho}_1^*$ and $\boldsymbol{\rho}_2^*$ are given by

$$\boldsymbol{\rho}^* = \begin{pmatrix} \boldsymbol{\rho}_1^* \\ \boldsymbol{\rho}_2^* \end{pmatrix}, \tag{5.3.5}$$

$$\dim(\boldsymbol{\rho}_i^*) = \dim(A_i), \quad i = 1,2. \tag{5.3.6}$$

Then it is easily shown that $\boldsymbol{\rho}'$ is the eigenvector of the following matrix A'(S) associating with the maximum eigenvalue.

$$A'(S) \triangleq \begin{pmatrix} A_1 & -A_3 \\ {}^t(-A_3) & A_2 \end{pmatrix}. \tag{5.3.7}$$

Also, it is clear that A'(S) is a positive matrix.

Namely, the optimum quantification vector can be derived from the eigenvector of a positive square matrix associating with the maximum eigenvalue, i.e., the Perron-Frobenius root, irrespective of V_1 or V_2 being empty or not.

The following theorem is due to Frobenius(cf. Karlin[1966]).

THEOREM(Frobenius) : If a positive square matrix A is strictly positive, i.e., all the entities of A are positive, or there exists an integer m(≥ 2) such that A^m is strictly positive, then A has a unique maximum eigenvalue λ^*, i.e., any other eigenvalue λ of A satisfies $|\lambda| < \lambda^*$, and the associated eigenvector $\mathbf{x}^*$ is unique and positive, i.e., any of its components is positive. Moreover, even if A is merely a positive matrix, there exists a positive eigenvector $\mathbf{x}^*$ associated with the maximum eigenvalue λ^*(of A).

From the above theorem, we can say that the optimum quantification vector $\rho^* = {}^t(\rho_1^*, \rho_2^*, \cdots, \rho_n^*)$ gives the cliques V_1 and V_2 of the structure theorem by the following rule:

$$\left.\begin{aligned} V_1 &= \{v_i \mid \rho_i > 0\} \\ V_2 &\quad \{v_i \mid \rho_i < 0\} \end{aligned}\right\} , \tag{5.3.8}$$

More precisely, the next theorem holds.

THEOREM 5.3.1 : If an s-graph S is connected, then the optimum quantification vector ρ^* is unique, and the sets V_1 and V_2 given by (5.3.8) coincide with the two cliques of the structure theorem. When S is disconnected, there exists an optimum quantification vector, i.e., an eigenvector of A(S) associating with the maximum eigenvalue, exists such that (5.3.8) gives the two cliques of the structure theorem.

PROOF : The theorem is easily verified by noting that ρ' is the eigenvector of A'(S) associating with the maximum eigenvalue and the relationships between ρ' and ρ^* and between A'(S) and A(S). Also, it should be noted that the condition of the connectedness on S is equivalent to the condition that there exists an integer m(≥ 1) such that all the entities of $A'(S)^m$ are positive. □

That is to say, when A(S) is regarded as a correlation matrix, then the principal component analysis of A(S) gives the aggregation of the s-graph S. However, for unbalanced groups, the above principal component analysis is superseded by centroid factor analysis as shown in the next section.

5.3.2 Aggregation of Unbalanced Group Structures into Tight Subgroups through a Quantification Method

In this section, the aggregation problem of unbalanced group structures into tight subgroups is discussed. As mentioned in section 5.1, this problem can be reduced to the derivation problem of the minimum balancing sets.

First of all, some characterization methods for the balancing and the minimum balancing sets of unbalanced groups are introduced in Section 5.3.2.1. In Section 5.3.2.2, an effective method for the derivation of the minimum

balancig sets is introduced by the use of the above characterizations.

5.3.2.1 Characterization of Balancing and the Minimum Balancing Sets by the Use of Sign Vectors

First of all, let us consider the characterization of balanced states by the use of sign vectors, where a *sign vector* $\mathbf{s}$ $(s_1, s_2, \cdots\cdots, s_n)$ is a vector whose element $s_i = +1$ or -1 for $i=1,2,\cdots,n$. The Structure Theorem by Cartwright and Harary has the following version(cf. Katai & Iwai[1978 a&c]):

LEMMA 5.3.1 : An s-graph S is balanced iff a sign vector $\mathbf{s} = {}^t(s_1, s_2, \cdots, s_n)$ exists, where n is the number of vertices of S, such that the adjacency matrix A(S) (a_{ij}) satisfies the following condition:

$$\text{if } a_{ij} \neq 0, \text{ then } a_{ij} = s_i s_j \quad \text{for } i,j = 1,2,\cdots,n, \tag{5.3.9}$$

in other words

$$a_{ij} s_i s_j \geq 0 \quad \text{for } i,j = 1,2,\cdots,n. \tag{5.3.9'}$$

PROOF : Let $V^+ = \{ v_i \in V \mid s_i = +1 \}$ and $V^- = \{ v_i \in V \mid s_i = -1 \}$. Then V^+ and V^- correspond to the subsets of the Structure Theorem. □

For example, the group in Fig. 5.3.1(a) is balanced. In this case, the following two sign vectors

$$s_1 = +1 \ , \ s_2 = -1 \ , \ s_3 = +1 \text{ or } s_1 = -1 \ , \ s_2 = +1 \ , \ s_3 = -1. \tag{5.3.10}$$

satisfies the following equations.

$$a_{12} = s_1 \cdot s_2 = -1 \ , \ a_{23} = s_2 \cdot s_3 = -1 \ , \ a_{31} = s_3 \cdot s_1 = -1 \tag{5.3.11}$$

On the other hand, the group in Fig. 5.3.1(b) is not balanced because the following equations

$$a_{12} = s_1 \cdot s_2 = -1 \ , \ a_{23} = s_2 \cdot s_3 = -1 \ , \ a_{31} = s_3 \cdot s_1 = -1 \tag{5.3.12}$$

have no solution (as a sign vector), for

$$s_1^2 \cdot s_2^2 \cdot s_3^2 = -1 \tag{5.3.13}$$

Each balancing set of a group has a corresponding balanced state derived from the original state by changing the signs of the relations contained in the balancing set.

By the above Lemma, the balancing set of a groups can be characterized as follows:

THEOREM 5.3.2 : A set L_b of the relations of a group S is a balancing set of S iff there exists a sign vector **s** such that the following conditions are satisfied.

$$v_i v_j \in L_b \quad \text{iff } a_{ij} \cdot s_i \cdot s_j < 0 \text{ for } i,j \quad 1,2,\ldots,n . \tag{5.3.14}$$

In the group of Fig. 5.3.2, the adjacency matrix A(S) is given as follows:

$$A(S) = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & -1 \\ 1 & 0 & -1 & -1 & 1 & -1 \\ 1 & -1 & 0 & -1 & -1 & 1 \\ 0 & -1 & -1 & 0 & -1 & 1 \\ 1 & 1 & -1 & -1 & 0 & -1 \\ -1 & -1 & 1 & 1 & -1 & 0 \end{pmatrix} \tag{5.3.15}$$

The group is unbalanced and one of the balancing sets is

$$L_b \quad \{v_1 v_3, v_3 v_4\} \tag{5.3.16}$$

The balanced state shown in Fig. 5.3.3 is derived by changing the relations $v_1 v_3$ and $v_3 v_4$. The sign vector **s** corresponding to the balancing set is

$$\mathbf{s} \quad (+1, +1, -1, -1, +1, -1) \tag{5.3.17}$$

That is, the set of members are split into two subgroups $\{v_1, v_2, v_5\}$ and $\{v_3, v_4, v_6\}$.

From theorem 5.3.2, the minimum balancing sets for an s-graph S are characterized by the next theorem.

THEOREM 5.3.3 : Let S be an s-graph and A(S) (a_{ij}) be its adjacency

matrix. Then a set L_b^* of lines of S is a minimum balancing set for S iff there exists a sign vector $\mathbf{s}^*$ such that

$$\hat{d}(\mathbf{s}^*) = \max_{\mathbf{s}:\text{sign vector}} \hat{d}(\mathbf{s}) \tag{5.3.18}$$

and

$$v_i v_j \in L_b^* \text{ iff } a_{ij} s_i^* s_j^* = -1 \text{ for } i,j = 1,2,\cdots,n, \tag{5.3.19}$$

where

$$\hat{d}(\mathbf{s}) \quad {}^t\mathbf{s}A(S)\mathbf{s} \quad ({}^t\mathbf{s}: \text{the tranposed vector of } \mathbf{s}). \tag{5.3.20}$$

PROOF : Let L_b be a balancing set of S and $c(L_b)$ be its number of elements. Then from theorem 5.3.1, $c(L_b)$ satisfies the following equation:

$$c(L_b) \quad \frac{1}{4}\{2m \quad \sum_{i,j=1}^{n} a_{ij} s_i s_j\} \quad \frac{1}{4}\{2m \quad \hat{d}(s)\} , \tag{5.3.21}$$

where m is the number of lines of S. A minimum balancing set is one whose $c(L_b)$ is minimum. □

5.3.2.2 Derivation of the Minimum Balancing Sets through Factor Analysis Techniques

From theorem 5.3.3, the minimal balancing sets of an s-graph S are given by the following quantification vectors (sign vectors) $\boldsymbol{\rho}$'s satisfying

$$Q \quad {}^t\boldsymbol{\rho}A(S)\boldsymbol{\rho} \rightarrow \max \tag{5.3.22}$$

subject to

$$\rho_i = +1 \text{ or } -1 \quad \text{for } i = 1,2,\cdots,n. \tag{5.3.23}$$

It is difficult, however, to treat the condition(5.3.23). Hence, we replace it by the following weaker condition:

$$\sum_{i=1}^{n} \rho_i^2 = n . \tag{5.3.24}$$

It is easily verified that the solution vector $\boldsymbol{\rho}^*$ of (5.3.22) subject to (5.3.24) is given as the eigenvector of A(S) which is associated with the maximum eigenvalue λ^* and also satisfies (5.3.24), i.e.,

$$A(S)\boldsymbol{\rho}^* = \lambda^*\boldsymbol{\rho}^* \quad , \tag{5.3.25}$$

$$\|\boldsymbol{\rho}^*\| = \sqrt{n} \quad . \tag{5.3.26}$$

Consequently, the quantification method for balanced group structure is supposed to have a tight connection with the determination of the optimum sign vectors $\mathbf{s}^*$'s, i.e., with the derivation of the minimum balancing sets.

Furthermore, we have the following estimate of the measure of degree of balance introduced by Abelson & Rosenberg by the use of the maximum eigenvalue λ^* of A(S). As aforementioned, the measure is given as $c(L_b^*)$, the number of elements in a minimum balancing set. From (5.3.21), the relation of $c(L_b^*)$ to λ^* is given as the following theorem (cf. Katai & Iwai[1978c]).

THEOREM 5.3.4 : The number $c(L_b^*)$, i.e., the minimum number of changes of relations necessary to restore balance, has the following lower bound:

$$c(L_b^*) \geq \left\lceil \frac{2m - \lambda^* n}{4} \right\rceil \quad , \tag{5.3.27}$$

where m and n are the numbers of the lines and the vertices of S, respectively, λ^* is the maximum eigenvalue of the symmetric matrix A(S), and $\lceil r \rceil$ is the minimum integer not less than r.

PROOF : It is obvious that

$$\hat{d}(\mathbf{s}) = {}^t\mathbf{s}A(S)\mathbf{s} \leq \max_{\|\boldsymbol{\rho}\|=\sqrt{n}} {}^t\boldsymbol{\rho}A(S)\boldsymbol{\rho} = \lambda^* n \quad , \tag{5.3.28}$$

where $\boldsymbol{\rho}$ is a real valued n-dimensional vector and $|\boldsymbol{\rho}|$ is its Euclidian norm. □

Let us consider the case as shown in Fig 5.3.4. The corresponding adjacency matrix A(S) is given by

$$A(S) = \begin{pmatrix} 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{pmatrix} , \tag{5.3.29}$$

and the solutions are as follows :

$$\lambda^* = 3.178 \quad , \tag{5.3.30}$$

$$\varphi^* = {}^t(1.27, -1.01, -1.17, 1.86, 1.23, -0.512, -0.111, 0.348, -0.727, 0.338) \quad . \tag{5.3.31}$$

As aforementioned, the optimum quantification vector φ^* attains the maximum value of Q subject to a weakened condition (5.3.24) of true condition (5.3.23). Hence, φ^* can be regarded as the ideal vector which attains the maximum value of $\hat{d}$ (cf.(5.3.20)). Let us consider the most proximate sign vector $\mathbf{s}'$ to the ideal vector $\boldsymbol{\rho}^*$. When the proximity is measured by the inner product of $\mathbf{s}$ and φ^*, then $\mathbf{s}'$ ${}^t(s'_1, s'_2, \cdots, s'_n)$ is given by (cf.(5.3.8))

$$s_i = \begin{cases} 1 & \text{if } \rho_i^* \geq 0 \\ -1 & \text{if } \rho_i^* < 0 \end{cases} \qquad \text{for } i = 1,2, \cdots ,n. \tag{5.3.32}$$

In the above example, we have

$$\mathbf{s}' \quad {}^t(1, -1, -1, 1, 1, -1, -1, 1, -1, 1,) \ . \tag{5.3.33}$$

The corresponding balancing set L'_b is follows(cf. Fig. 5.3.5):

$$L'_b \quad \{v_2v_3, v_5v_7, v_6v_8\} \ . \tag{5.3.34}$$

From theorem 5.3.4, we have

$$c(L_b^*) \geq \left\lceil \frac{1}{4}(38 \quad 3.178 \times 10) \right\rceil = \lceil 1.555 \rceil = 2 \quad . \tag{5.3.35}$$

The balancing set L_b' yields $c(L_b') = 3$, hence it may not be a minimum balancing set. However, in the graph S of Fig. 5.3.4, there are three cycles $\sigma_1 = v_1v_2, v_2v_3, v_3v_1$, $\sigma_2 = v_3v_5, v_5v_7, v_7v_8, v_8v_3$, and $\sigma_3 \quad v_6v_7, v_7v_9, v_9v_{10}$, $v_{10}v_8, v_8v_6$ having negative signs such that no line is common in any pair of the cycles. Therefore, we have $c(L_b') \geq 3$. Hence the above balancing set L_b' is a minimum balancing set for S in Fig. 5.3.4. That is to say, the modified method of quantification method IV directly gives the minimum balancing set.

In general, we cannot expect the ideal case such as the above; the optimum quantification vector $\boldsymbol{\rho}^*$ only specifies the neighbourhood of the optimum sign vector $\mathbf{s}^*$. On the other hand, the rule (5.3.18) for the minimum balancing sets is directly related to the so-called centroid method in factor analysis techniques. In fact, the first factor loading vector $\mathbf{f}^* = {}^t(f_1^*, f_2^*, \cdots, f_n^*)$ for the matrix A(S) (when considered as a correlation matrix) is paraphrased by the following equation (for details, see Horst [1965]):

$$\mathbf{f}^* = (\sqrt{{}^t\mathbf{s}^*A(S)\mathbf{s}^*}\,)^{-1}\mathbf{g}^* \quad , \tag{5.3.36}$$

where the vector $\mathbf{g}^* = {}^t(g_1^*, g_2^*, \cdots, g_n^*)$ is defined by

$$\mathbf{g}^* = A(S)\mathbf{s}^* \quad , \tag{5.3.37}$$

and the optimum sign vector $\mathbf{s}^*$ is in turn given by

$$s_i^* = \begin{cases} 1 & \text{if } g_i^* \geq 0 \ , \\ & \qquad \text{for } i \quad 1,2,\cdots,n. \\ -1 & \text{if } g_i^* < 0, \end{cases} \tag{5.3.38}$$

However, the diagonal elements of A(S) are zero, so we cannot regard A(S) as a correlation matrix. Yet we can apply a practical procedure used to search for the optimum sign vector $\mathbf{s}^*$ in the centroid method to this area also. The procedure is based on the reciprocal relations (5.3.37) and (5.3.38) between the two vectors $\mathbf{s}^*$ and $\mathbf{g}^*$. But the solution of these two relations is not necessarily unique, so the procedure may produce not the optimum solution but a suboptimum one. (Unfortunately, there is no efficient way to determine $\mathbf{s}^*$ in general cases, the only way being to list all the n-dimentional sign vectors and to search methodically.)

The procedure is recursively defined as follows (for details, see Horst [1965]): Let $\mathbf{s}^i = {}^t(s_1^i, s_2^i, \cdots, s_n^i)$ and $\mathbf{g}^i = {}^t(g_1^i, g_2^i, \cdots, g_n^i)$ be the i^{th} approximation of $\mathbf{s}^*$ and $\mathbf{g}^*$, respectively. Then the $i+1^{th}$ approximation $\mathbf{s}^{i+1}$ and $\mathbf{g}^{i+1}$ are given by

$$s_j^{i+1} \begin{cases} s_j^i & \text{if } j \neq \alpha_i \\ -s_{\alpha_i}^i & \text{if } j = \alpha_i \end{cases}, \tag{5.3.39}$$

$$\mathbf{g}^{i+1} = A(S)\mathbf{s}^{i+1} , \tag{5.3.40}$$

where α_i is the position specified by

$$d_{\alpha_i}(\mathbf{s}^i) = \min_j d_j(\mathbf{s}^i) < 0 \tag{5.3.41}$$

and $d_j(\mathbf{s}^i)$ is defined by

$$d_j(\mathbf{s}^i) = s_j^i \cdot g_j^i = s_j^i \sum_{k=1}^{n} a_{jk} s_k^i . \tag{5.3.42}$$

The procedure starts with arbitrarily selected initial sign vector $\mathbf{s}^0$ and $\mathbf{g}^0 = A(S)\mathbf{s}^0$. If the conditions

$$d_j(\mathbf{s}^k) \geq 0 \quad \text{for all } j = 1,2, \cdots, n \tag{5.3.43}$$

are satisfied for some k, then the procedure stops and the resultant solution is $\mathbf{s}^k$(and $\mathbf{g}^k$). In the usual factor analysis problem, A(S) corresponds to a real valued correlation matrix, so the position α_i specified by (5.3.41) is uniquely determined. However, in our case, the adjacency matrix A(S) is integer valued, the uniqueness of α_i being sometimes violated. In such cases, the position α_i is arbitrarily assigned to one of the candidates.

When the members (vertices) of the group (graph S) are divided into two subgroups according to the sign vector $\mathbf{s}^i$, the value $d_j(\mathbf{s}^i)$ defined by (5.3.42) is regarded as a degree of consistency of the grouping with respect to the j^{th} member. The degree is minimum (and negative) for $\alpha_i{}^{th}$ member. The member changes his group membership according to the rule (5.3.39). Therefore the process means the exchange of the member having the strongest strain at

that moment from one group to the other, so as to increase his degree of consistency (cf.Katai, Iwai et al.[1975a]), i.e.,

$$d_{\alpha_i}(\mathbf{s}^i) < d_{\alpha_i}(\mathbf{s}^{i+1}) \quad . \tag{5.3.44}$$

The relation of the degrees with (5.3.20) is stated by

$$\hat{d}(\mathbf{s}) = {}^t\mathbf{s}^i A(S)\mathbf{s}^i = \sum_{j=1}^{n} d_j(\mathbf{s}^i) \quad . \tag{5.3.45}$$

From (5.3.39) and (5.3.41), the above sum strictly increases with i and is bounded by the value $\hat{d}(\mathbf{s}^*)$. Therefore the process always stops (by the condition (5.3.43)) and the final grouping $\mathbf{s}^k$ satisfies the reciprocal relations (5.3.37) and (5.3.38).

As aforementioned, we cannot expect that the above $\mathbf{s}^k$ gives a minimum balancing set L_b^*. Therefore, the procedure must be repeated using initial sign vectors which do not appear in previous steps. Also, to avoid needless repetitions, if the succeeding procedure yields a sign vector $\mathbf{s}$ such that $\mathbf{s}$ or $-\mathbf{s}$ appeared in previous steps, then the procedure stops and a new initial vector $\mathbf{s}^0$ is selected and a new trial starts. Indeed, the selection of initial vectors plays an important role in the efficiency of the procedure. In the last part of this section, we propose a method for the selection.

The above procedure is directly related to the method introduced by Abelson & Rosenberg (1958) used to search for the number $c(L_b^*)$. Their method is also based on the adjacency matrices and employs certain rewriting rules (transformations) directly acting on the adjacency matrices. On the other hand, in our procedure, the sign vectors play the same role as the transformations of adjacency matrices and the adjacency matrices are not rewritten, so that the procedure becomes much simpler.

We illustrate the procedure using an example in Fig. 5.3.4. Table 5.3.1 shows the procedure applied to the above example. We set the first initial sign vector as $\mathbf{s}^0 = {}^t(1, 1, \cdots, 1)$. The second column represents $d_j(\mathbf{s}^0)$ for $j = 1, 2, \cdots, 10$. The encircled numeral at j 3 indicates the minimum of the above values. Hence, the third component of $\mathbf{s}^0$ is changed to -1 yielding $\mathbf{s}^1$. Similar operations continue until $\mathbf{s}^3$ is produced which satisfies the condition (5.3.43) and gives $\hat{d}(\mathbf{s}^3) = 22$ and $c(L_b) = 4$. The next initial sign vector $\mathbf{s}^0 (= {}^t(-1, 1, \cdots, 1))$ does not appear in previous

steps. The final vector $\mathbf{s}^2$ of this trial gives $\hat{d}(\mathbf{s}^2)$ 18 and $c(L_b) = 5$. In the third trial, the sign vector $\mathbf{s}^1$ coincides with the sign vector $\mathbf{s}^2$ of the first trial, hence the procedure enters the next trial. At seventh trial, vector $\mathbf{s}^4$ yields that $\hat{d}(\mathbf{s}^4) = 26$ and $c(L_b) = 3$. The corresponding balancing set is the same as L_b' given by (5.3.34).

By continuing the above procedure, it is shown that there is no other minimum balancing set, i.e., the balancing set $\{v_2v_3, v_5v_7, v_6v_8\}$ is the unique minimum balancing set. Concerning the setting of the initial sign vectors, we take note of the aforementioned fact that the quantification vector $\boldsymbol{\rho}^*$ specifies the neighbourhood of the optimum sign vector $\mathbf{s}^*$, i.e., $\mathbf{s}^*$ is supposed to be located in the neighbourhood of $\boldsymbol{\rho}^*$. Hence, we can expect that the procedure will reach the optimum sign vectors much faster using sign vectors which are proximate to the ideal vector $\boldsymbol{\rho}^*$ as the initial sign vectors, instead of using the vectors (1, 1, $\cdots$, 1), (-1, 1, $\cdots$,1) etc. which have no relation with the adjacency matrix A(S). In the above example, as aforementioned, the most proximate sign vector $\mathbf{s}'$ to $\boldsymbol{\rho}^*$ directly gives the unique minimum balancing set L_b' (= L_b^*) represented by (5.3.34).

5.3.3 Aggregation of Unbalanced Group Structures through Graph Theoretical Methods

The notion of structural balance is originally graph theoretical. Hence some graph theoretical methods are supposed to offer more effective algorithms to derive the minimum balancing sets compared to the algorithm in Section 5.3.2. In Section 5.3.3.1, we consider the simplification of the aggregation problem by the use of the block decomposition of group structures. In Section 5.3.3.2, we consider an effective algorithm to derive the minimum balancing sets for unbalanced groups with planar graph structures, by introducing the notion of dual graph.

First of all, the following notions in graph theory are necessary in developing our theory, which are mainly from Harary(1969) and Ore(1962).

A *graph G* is a pair (L,V) : V is a finite, nonempty set of *vertices* (*points*) and L is a set of unordered pairs of distinct vertices. G can be identified with the subset L of $V \otimes V$, the set of unordered pairs of distinct elememnts of V. Each pair x = {u,v} (x=uv) of vertices is called a *line* of G and x is said to join u and v. The vertices u and v are said to be *adjacent* and the vertex u and the *line* x are said to be *incident* with each

other. A connected sequence of lines is said to be a *path* when no vertices appear more than once. A closed path is called a *circuit*. A *signed graph* (*s-graph*) S is a graph such that each line has a positive value " + " or a negative value " − " attached. We sometimes denote + and − by " +1 " and " -1 ", respectively. We denote by G(S) the graph G obtained from an s-graph S by eliminating the signs of lines of S.

The following notions for a graph G are also necessary. For an s-graph S, these notions are defined in the same manner as for G(S), by reserving the signs of lines in S, except for the notion of *dual graph*. In the next section, we introduce the notion of the dual graph of an s-graph.

A *subgroup* of G is a graph having all of its vertices and lines in G. The *removal of a point v* from a graph G reduces it to a graph G − v which is the maximal subgraph of G not containing v. A graph is *connected* iff every pair of points are joined by a path. A maximal connected subgraph of G is called a *connected component* or simply a *component* of G. A *cutpoint* of a graph is a vertex whose removal increases the number of connected components. A graph which is connected, nontrivial (i.e., contains at least two points), and has no cut points is called *nonseparable*. A *block component* of a graph is a maximal nonseparable subgraph. If G is nonseparable, then G itself is often called a *block*. Any separable graph G can be decomposed into block components and any circuit of G is contained in exactly one of them, for example Fig. 5.3.6. A *spanning subgraph* of G is a subgraph containing all the vertices of G. A *spanning tree* of G is a spanning subgraph which is connected and contains no circuits.

Another notion concerning graphs is *cycle* defined in a rather algebraic way as follows. We introduce a vector space over the two element field $F_2 = \{0,1\}$, in which $1 + 1 = 0$. A 1-*chain* of a graph G is a formal linear combination $\sum_i \varepsilon_i x_i$ of lines of G, and 0-*chain* is a sum $\sum_i \varepsilon_i v_i$ of vertices where each ε_i is an element of F_2. The boundary operator ∂ transforms 1-chains to 0-chains according to the rules:

(1) $\partial(x + x') = \partial x + \partial x'$,

(2) if $x = uv$, then $\partial x = u + v$.

In Fig. 5.3.7, the 1-chain $\sigma = x_1 + x_2 + x_4 + x_5$ has boundary

$$\partial\sigma = (v_1 + v_2) + (v_2 + v_3) + (v_1 + v_4) + (v_2 + v_4)$$

$= v_2 + v_3$.

A 1-chain with boundary 0 is called a *cycle* (or *cycle vector*) of G and can be regarded as a set of line-disjoint circuits. The collection of all cycle vectors forms a vector space over F_2 called the *cycle space* of G. A *cycle basis* (or a *set of fundamental cycles*) of G is defined as a basis for the cycle space of G.

A graph is *planar* iff it can be embedded in a plane with vertices represented by points, edges represented by lines, so that lines intersect only at points representing vertices : a *plane* graph is a planar graph which has already been embedded in a plane. The regions defined by a plane graph are referred to as its *faces* and the unbounded region is called the *exterior face*. Given a plane graph G, its *dual graph* D(G) is constructed as follows. Place a vertex in each face of G (including the exterior face) and, if two faces have a line x in common, join the corresponding vertices by a line $\hat{x}$ crossing only x (see Fig. 5.3.8). Each line x of G corresponds to a line $\hat{x}$ of D(G) denoted by d(x) and also each line x of D(G) corresponds to a line of G denoted by $d^{-1}(x)$. If G is nonseparable, then D(G) has no loops but may have multiple lines.

5.3.3.1 Simplification of the Problem by the Use of Block Decomposition of Graph Structures of Groups

The following lemma is beneficial for considering the aggregation problem.

LEMMA 5.3.2 : An s-graph S is balanced iff every block component in it is balanced.

PROOF : From the definition of block component, every circuit of S is contained in exactly one of its block components. So the lemma is a direct consequence of the definition of balanced states. □

The lemma leads to the next theorem(cf. Katai & Iwai[1978a]).

THEOREM 5.3.5 : Let S be an s-graph and B_1, B_2, ... , and B_t be its block components. Then a set L of lines of S is a balancing set for S iff L is represented as follows :

$$L = \bigcup_{i=1}^{t} L_i \quad , \tag{5.3.46}$$

where L_i is a balancing set for B_i for $i = 1, 2, \cdots, t$. A set L is a minimum balancing set iff L_i is a minimum balancing set for B_i for $i = 1, 2, \cdots, t$.

PROOF : Noting that any line of S is contained in exactly one of B_1, B_2, $\cdots$, B_k, the theorem is easily verified using the above lemma. □

According to the above fact, the characterizations of the balancing sets or the minimum balancing sets for an s-graph are reduced to those for each of its block components. For example, in Fig. 5.3.9, the balancing sets or the minimum balancing sets for S is the union of the balancing sets or the minimum balancing sets for B_1, B_2, B_3, and B_4, respectively. An s-graph without circuits is always balanced (from the definition), therefore B_2 is balanced. B_4 has already been discussed. B_1 and B_3 are planar graphs. We will discuss them in later sections. Hence, we can assume that given s-graphs are nonseparable.

When S is a planar s-graph, the following graph theoretical method is quite effective for the derivation of the minimum balancing set for S.

5.3.3.2 Aggregation of Unbalanced Groups Prescribed by Planar Graph Structures

As mentioned previously, a planar graph can be represented by a plane graph. So we assume that given nonseparable s-graph are plane graphs. First of all, we introduce the notion of dual graph for plane s-graph as follows (cf. Katai & Iwai [1978a & c]).

DEFINITION 5.3.1 : Let S be a plane s-graph. Then the *dual graph* D(S) of S is the graph D(G(S)) together with the sign of each vertex which is defined as follows. If a vertex of D(G(S)) corresponds to a (inner) face of S, then its sign is that of the circuit which encircle the face. If it corresponds to the exterior face, then the sign is that of the circumferential circuit (cycle), i.e., the circuit which encloses S.

Fig. 5.3.10 shows an example (cf. Fig. 5.3.8).

The balanced states of plane s-graph are characterized as follows :

LEMMA 5.3.3 : A plane s-graph is balanced iff every vertex of D(S) has a positive sign.

PROOF : The lemma is easily verified from the following facts. According to the definitions of cycle and cycle basis, a circuit of a graph S is also its cycle, and the circuits corresponding to (inner) faces constitute a cycle basis of the cycle space of S. That is, every circuit (cycle) of S is represented by a sum of some of them. According to the following lemma 5.3.4, a circuit represented by a sum of positive circuits is also positive. Therefore the condition that every circuit has a positive sign means that the circuits which constitute a cycle basis have positive signs. □

LEMMA 5.3.4 : Let S be an arbitrary s-graph and σ_i be a cycle (circuit) of S for i = 1, 2, ··· ,k. Then the following equality holds:

$$s(\sigma_1 + \sigma_2 + \cdots + \sigma_k) = s(\sigma_1) \cdot s(\sigma_2) \cdots s(\sigma_k) , \qquad (5.3.47)$$

where $s(\sigma)$ is the sign of cycle (circuit) σ.

PROOF : If the case of k 2 is verified, then general cases are easily verified using induction on the number of cycles k. If $\sigma_i = \sum_{j=1}^{n} \varepsilon_j^i x_j$ for i 1,2, then $\sigma_1 + \sigma_2 = \sum_{j=i}^{n} (\varepsilon_j^1 + \varepsilon_j^2) x_j$. Let $s(x_j)$ denote the sign of line x_j of S. Then $s(\sigma_i) = \prod_{x_j \in L_i} s(x_j)$ for i 1, 2, where L_i is the set of lines x_j contained in σ_i, i.e., $\varepsilon_j^i = 1$. Similarly, $s(\sigma_1 + \sigma_2) \prod_{x_j \in L_1 \cup L_2 - L_1 \cap L_2} s(x_j)$ but this is the same as $\prod_{x_j \in L_1} s(x_j) \cdot \prod_{x_j \in L_2} s(x_j) = s(\sigma_1) \cdot s(\sigma_2)$ because of the fact that $(-1)^2 = 1^2 = 1$. □

The above lemma also leads to another consequence as follows:

LEMMA 5.3.5 : Let S be a plane s-graph. Then the number n_u of vertices with negative sign in the graph D(S) is even.

PROOF : The vertex v_o of D(S) placed in the exterior face of S corresponds to the circumferential circuit σ_o of S. That is, the sign of the vertex is defined by that of σ_o. Let σ_1, σ_2, ··· , and σ_m be the circuits of S corresponding to the inner faces of S. By the defintion of plane graph,

$$\sigma_o = \sigma_1 + \sigma_2 + \cdots + \sigma_m \quad . \qquad (5.3.48)$$

According to lemma 5.3.4, the sign of v_0 equals the product of the signs of the inner vertices of D(S). Therefore, the product of signs of all the ver-

tices of D(S) equals one. □

Let $L = \{x_1, x_2, \cdots, x_n\}$ be a set of lines of a graph. We denote by $\sigma_1(L)$ the 1-chain defined as $\sum_{i=1}^{n} x_i$ and by d(L) the set of lines of its dual graph defined as $\{d(x_1), d(x_2), \cdots, d(x_n)\}$. If $V = \{v_1, v_2, \cdots, v_m\}$ is a set of vertices of a graph, then $\sigma_0(V)$ is the 0-chain $\sum_{i=1}^{m} v_i$. Using the above notations, we can verify the following equality.

LEMMA 5.3.6 : For a given s-graph S, let L be the set of lines of S whose signs are changed. Then, in the graph D(S), the set of vertices V(L) whose signs are changed is represented as follows :

$$\sigma_0(V(L)) = \partial(\sigma_1(d(L))), \tag{5.3.49}$$

where ∂ is the boundary operator.

PROOF : We use induction on the number n of elements of the set $L = \{x_1, x_2, \cdots, X_n\}$. Let n equal 1, i.e., $L = \{x_1\}$. Then the set V(L) is equal to $\{v_1, v_1'\}$ where v_1 and v_1' are the vertices of D(S) corresponding to the faces of S incident to the line x_1, i.e., $d(x_1) = v_1v_1'$. Then we have

$$\sigma_0(V(L)) = \sigma_0(\{v_1, v_1'\}) = v_1 + v_1', \tag{5.3.50}$$

$$\partial(\sigma_1(d(L))) = \partial(\sigma_1(\{v_1v_1'\})) = \partial(v_1v_1') = v_1 + v_1'. \tag{5.3.51}$$

So the lemma is verified.

If the case n = k is verified, then the case n = k + 1 is proved as follows. In this case

$$L = \{x_1, x_2, \ldots, x_k, x_{k+1}\} = L_0 \cup \{x_{k+1}\}, \tag{5.3.52}$$

where $L_0 = \{x_1, x_2, \ldots, x_k\}$. According to the definition of σ_1 and d,

$$\sigma_1(d(L)) = \sigma_1(d(L_0)) + d(x_{k+1}). \tag{5.3.53}$$

Suppose that we first change the signs of lines contained in L_0 and then change the sign of the line x_{k+1}. Then the resultant signs of the vertices of D(S) are the same as those obtained by changing the signs of lines in L simultaneously. Let $d(x_{k+1})$ equal $v_{k+1}V_{k+1}'$. If $v_{k+1}(v_{k+1}')$ is contained in $V(L_0)$, then its sign is changed twice under the two succeeding sign changes in the above supposition, i.e., $v_{k+1}(v_{k+1}')$ has its original sign and is not contained in V(L). If $v_{k+1}(v_{k+1}')$ is not contained in $V(L_0)$, then $v_{k+1}(v_{k+1}')$ is contained in V(L).

Therefore, from the rule that $1 + 1 \quad 0$ on the field $F_2 = \{0,1\}$, the following equality holds:

$$\sigma_0(V(L)) = \sigma_0(V(L_0)) + \partial(d(x_{k+1})) \quad . \tag{5.3.54}$$

By the induction assumption

$$\sigma_0(V(L_0)) = \partial(\sigma_1(d(L_0))) \quad , \tag{5.3.55}$$

we have

$$\begin{aligned} \sigma_0(V(L)) &= \partial(\sigma_1(d(L_0))) + \partial(d(x_{k+1})) \\ &= \partial(\sigma_1(d(L_0)) + d(x_{k+1})) \\ &= \partial(\sigma_1(d(L))) \quad . \end{aligned} \tag{5.3.56}$$

So the case of $n = k + 1$ is verified. □

The next lemma characterizes the balancing sets for arbitrarily given plane s-graphs.

LEMMA 5.3.7 : Let S be a plane s-graph and V_u be the set of vertices of D(S) with negative sign. Then a set L_b of lines of S is a balancing set of S iff the following formula is valid:

$$\sigma_0(V_u) = \partial(\sigma_1(d(L_b))) \quad . \tag{5.3.57}$$

PROOF : The lemma is a direct consequence of lemma 5.3.3 and lemma 5.3.6. □

The one-to-one correspondence between the lines of a plane graph and those of its dual graph is also true in the case of plane s-graphs. That is, if $d(L_b)$ is given by a set L of lines of D(S), then L_b equals $d^{-1}(L)$. According to the definition of boundary operator and the fact indicated by lemma 5.3.5, the above lemma is also stated as follows (cf. Katai & Iwai[1978c]):

THEOREM 5.3.6 : Let S be a plane s-graph and $V_u = \{v_1, v_2, \cdots, v_{2m}\}$ be the set of negative vertices of D(S). Then L_b is a balancing set for S iff there exist sets $L_1, L_2, \cdots, L_m$ and L_{m+1} of lines of D(S) and sets (pairs) V_1 $\{v_{p_1}, v_{p_1}'\}$, V_2 $\{v_{p_2}, v_{p_2}'\}, \cdots$, and $V_m = \{v_{p_m}, v_{p_m}'\}$ of vertices of D(S)

such that

$$L_b \quad d^{-1}(L_1 \cup L_2 \cup \ldots \quad L_m \cup L_{m+1}), \tag{5.3.58}$$

$$L_i \cap L_j = \Phi \text{ for } i, j \quad 1, 2, \ldots, m+1, i \neq j, \tag{5.3.59}$$

$$V_1 \cup V_2 \cup \ldots \cup V_m = V_u, \tag{5.3.60}$$

$$\text{therefore, } V_i \cap V_j = \Phi \text{ for } i, j = 1, 2, \ldots, m, i \neq j, \tag{5.3.61}$$

L_i $(i = 1, 2, \ldots, m)$: the set of lines contained in a path joining the vertices V_{p_i} and $V_{p_i'}$, (5.3.62)

L_{m+1} : the set of lines contained in a cycle of D(S) (including a 0 cycle, i.e., L_{m+1} may be empty). (5.3.63)

PROOF : The first half of this equivalence is easily verified as follows. From equations (5.3.58) and (5.3.59) and the linearity of boundary operator ∂, we have

$$\partial(\sigma_1(d(L_b))) \quad \sum_{i=1}^{m+1} \partial(\sigma_1(L_i)). \tag{5.3.64}$$

According to the conditions (5.3.62) and (5.3.63)

$$\partial(\sigma_1(L_i)) \quad v_{p_i} + v_{p_i'} \text{ for } i = 1, 2, \ldots, m, \tag{5.3.65}$$

$$\partial(\sigma_1(L_{m+1})) = 0. \tag{5.3.66}$$

Using the above three equations,

$$\partial(\sigma_1(d(L_b))) = v_{p_1} + v_{p_1'} + v_{p_2} + v_{p_2'} + \ldots + v_{p_m} + v_{p_m'}. \tag{5.3.67}$$

By the condition (5.3.60)(and (5.3.61)),the above equations lead to

$$\partial(\sigma_1(d(L_b))) = \sigma_0(V_u). \tag{5.3.68}$$

The other half of the equivalence shall be proved by induction on m. Suppose that m = 1, i.e., $V_u = \{v_1, v_2\}$. Let $D(L_b)$ be the subgraph of D(S) containing exactly the lines of $d(L_b)$. It is clear that v_1 and v_2 are contained in $D(L_b)$. Let B_1, B_2, $\cdots$, and B_t be the connected components of $D(L_b)$ and $L(B_i)$ be the set of lines contained in B_i for i = 1, 2,$\cdots$, t.

Then it is obvious that

$$d(L_b) = \bigcup_{i=1}^{t} L(B_i), \tag{5.3.69}$$

$$B_i \cap B_j \quad \Phi \text{ for } i, j = 1, 2, \ldots, t, i \neq j, \tag{5.3.70}$$

$$\partial(\sigma_1(d(L_b))) = \sum_{i=1}^{t} \partial(\sigma_1(L(B_i))) = \sigma_0(V_u) \quad v_1 + v_2. \tag{5.3.71}$$

From (5.3.70), (5.3.71) and the fact that every boundary contains an even number of vertices, it is easy to see that the vertices v_1 and v_2 are contained in the same connected component. Without loss of generality, we can assume that

$$v_1, v_2 \in B_1 . \tag{5.3.72}$$

That is, there is a path in B_1 joining v_1 and v_2. Let L_1 be the set of lines contained in the path. Then, by (5.3.65) (5.3.71), we have

$$\partial(\sigma_1(L(B_1) \quad L_1)) = 0, \tag{5.3.73}$$

$$\partial(\sigma_1(L(B_i))) = 0 \text{ for } i \quad 2, 3, \ldots, t . \tag{5.3.74}$$

Therefore the set L_1 and the set L_2 defined by

$$L_2 \quad d(L_b) \quad L_1 = (L(B_1) \quad L_1) \cup L(B_2) \cup \ldots \cup L(B_t) \tag{5.3.75}$$

satisfy the conditions of the theorem.

We prove the case m k + 1 assuming that the theorem is true for the case m = k as follows. Using a method similar to that for the derivation of (5.3.72), we can assume that two vertices v_{p_1} and $v_{p_1'}$ of V_u are contained in a connected component of $D(L_b)$. Let L_1 be the set of lines defined similarly as above and V_1 be the set $\{v_{p_1}, v_{p_1'}\}$. Then we have

$$\partial(d(L_b) \quad L_1) = \sigma_0(V_u \quad V_1). \tag{5.3.76}$$

By the induction hypothesis, there exist sets $L_2, L_3, \cdots$ and L_{k+2} and $V_2, V_3, \cdots$ and V_{k+1} satisfying the conditions (5.3.59), (5.3.62) and (5.3.63) (where m k) together with the following conditions:

$$\bigcup_{i=2}^{k+2} L_i = d(L_b) \quad L_1, \tag{5.3.77}$$

$$\bigcup_{i=2}^{k+1} V_i = V_u - V_1. \tag{5.3.78}$$

Therefore, the sets L_1, L_2, $\cdots$ and L_{k+2} and also V_1, V_2, $\cdots$ and V_{k+1} satisfy the condition of the theorem. □

The above theorem says that to accomplish a balanced state, each unbalanced vertex must be connected to another unbalanced vertex by a path. In theorem 5.3.5, to be a minimum balancing process, the set L_i must be composed of lines contained in a path of minimum length joining the vertices v_{p_i} and $v_{p_i'}$, for i = 1, 2, $\cdots$ and m. Also, L_{m+1} must be empty. Moreover, the pairing V_1 V_2, $\cdots$ and V_m must be chosen such that the total sum of the above minimum length is also minimum. These statements are formally described by the following theorem:

THEOREM 5.3.7 : Let S and V_u be the same as in theorem 5.3.6. Then L_b^* is a minimum balancing set for S iff there exist sets L_1^*, L_2^*, $\cdots$, and L_m^* of lines of D(S) and sets $V_1^* = \{v_{p_1}, v_{p_1'}\}$, $V_2^* = \{v_{p_2}, v_{p_2'}\}$, $\cdots$, and V_m^* $\{v_{p_m}, v_{p_m'}\}$ of vertices of D(S) satisfying condition (5.3.60) together with the conditions

$$L_b^* = d^{-1}(L_1^* \cup L_2^* \cup \ldots \cup L_m^*), \tag{5.3.58'}$$

L_i^* (i = 1, 2, ... , m) : the set of lines contained in a path of

minimum length joining the vertices v_{p_i} and $v_{p_i'}$, (5.3.62')

the pairs $V_1^*, V_2^*, \ldots, V_m^*$ are chosen such that the index $c(V_1, V_2, \ldots, V_m) = \sum_{i=1}^{m} c(V_i)$ is minimum, i.e.,

$$c(V_1^*, V_2^*, \ldots, V_m^*) = \min_{\substack{V_1, V_2, \ldots, V_m \\ V_1 \cup V_2 \cup \ldots \cup V_m = V_u}} c(V_1, V_2, \ldots, V_m), \tag{5.3.79}$$

where $c(V_i)$ is the length of the shortest path joining v_{p_i} and $v_{p_i'}$.

PROOF : We only have to show that if equality (5.3.79) holds, then condition (5.3.59) is satisfied. For example, the condition that $L_1^* \cap L_2^* = \Phi$ is proved as follows :

According to condition (5.3.62'),

$$\partial(\sigma_1(L_i^*)) = v_{p_i} + v_{p_i'}, \quad \text{for } i = 1, 2. \tag{5.3.80}$$

By the condition of the theorem that $V_1^* \cap V_2^* = \Phi$,

$$\partial(\sigma_1(L_1^*)) + \partial(\sigma_1(L_2^*)) \quad v_{p_1} + v_{p_1'} + v_{p_2} + v_{p_2'} . \tag{5.3.81}$$

Using the linearity of ∂,

$$\partial(\sigma_1(L_1^*) + \sigma_1(L_2^*)) \quad v_{p_1} + v_{p_1'} + v_{p_2} + v_{p_2'} . \tag{5.3.82}$$

Assume that

$$L_1^* \cap L_2^* = L_{12} \neq \Phi. \tag{5.3.83}$$

Let a set of lines $\tilde{L}$ be as follows :

$$\tilde{L} - L_1^* \cup L_2^* \quad L_{12} \subsetneq L_1^* \cup L_2^*. \tag{5.3.84}$$

Then, from the definition of " + " and (5.3.82), we have

$$\partial(\sigma_1(\tilde{L})) = \partial(\sigma_1(L_1^*) + \sigma_1(L_2^*)) = \sigma_0(\{v_{p_1}, v_{p_1'}, v_{p_2}, v_{p_2'}\}). \tag{5.3.85}$$

Hence, from lemma 5.3.7 and theorem 5.3.6, there exist sets of lines $\bar{L}_1$, $\bar{L}_2$ and $\bar{L}_3$, and also sets of vertices $\bar{V}_1 = \{\bar{v}_1, \bar{v}_1'\}$ and $\bar{V}_2 \ \{\bar{v}_2, \bar{v}_2'\}$ such that

$$\tilde{L} = \bar{L}_1 \cup \bar{L}_2 \cup \bar{L}_3 , \tag{5.3.86}$$

$\bar{L}_i$: the set of lines contained in a path joining $\bar{v}_i$ and $\bar{v}_i'$

for $i = 1, 2$, (5.3.87)

$\bar{L}_3$: the set of lines contained in a cycle, (5.3.88)

$$\bar{V}_1 \cup \bar{V}_2 = V_1^* \cup V_2^* = \{v_{p_1}, v_{p_1'}, v_{p_2}, v_{p_2'}\}. \tag{5.3.89}$$

It is obvious that

$$c(V_1^*) + c(v_2^*) = \#L_1^* + \#L_2^* > \#\tilde{L} \geq \#\bar{L}_1 + \#\bar{L}_2 \geq c(\bar{V}_1) + c(\bar{V}_2). \tag{5.3.90}$$

The above inequality and (5.3.89) contradict the optimality condition (5.3.79). Hence, we obtain $L_1^* \cap L_2^* = \Phi$. □

For the groups in Fig. 5.2.2(a), the negative vertices are v_1 and v_2 in Fig. 5.3.11. Therefore, the pairing is only $V_1 = \{v_1, v_2\}$. The paths of the minimum length connecting v_1 and v_2 are $\{v_1v_4, v_4v_2\}$ and $\{v_1v_6, v_6v_2\}$. The first one corresponds to case (b) in Fig. 5.2.2. Also, the second one

corresponds to case (d). Case (c) does not correspond to a minimum balancing process because it corresponds to the path $\{v_1v_3, v_3v_5, v_5v_2\}$(cf. Katai & Iwai, et al. [1975a] and Katai & Iwai[1978c]).

For seeking the minimum balancing sets for arbitrarily given s-graph S, theorem 5.3.7 gives the following algorithm.

Step 1 : Calculate the signs of circuits corresponding to the faces of S.

Step 2 : Make the dual graph D(S).

Step 3 : Choose the negative vertices and calculate the distances between all the pairs of negative vertices, where the distance of two vertices is the length of the shortest path joining them.

Step 4 : Make the complete graph whose lines are attached by the calculated distances.

Step 5 : Seek the optimum pairs V_1^*, V_2^*, $\cdots$ and V_m^* such that (5.3.60) and the criterion (5.3.79) hold. (In the definition of index $c(V_1^*, V_2^*, \cdots, V_m^*)$, $\#L_i^*$ equals the calculated distance beween v_{p_i} and $v_{p_i'}$, for i = 1, 2, $\cdots$, m).

Step 6 : Find the shortest path L_i^* joining the vertices of V_i^* for i = 1, 2, ... , m (cf. (5.3.62')).

Step 7 : Derive the minimum balancing set L_b^* according to the rule (5.3.58'), i.e.,

$$L_b^* = d^{-1}(L_1^* \cup L_2^* \cup \ \ldots \cup L_m^*).$$

In the operations of step 5 and 6, the optimum pairs V_1^*, V_2^*, $\cdots$ and V_m^* and paths L_1^*, L_2^*, $\cdots$and L_m^* are not necessarily unique. So the number of L_b^* produced by the last step may be more than one. As stated in theorem 5.3.7, a minimum balancing set L_b^* can be derived by the above algorithm, and a set L_b^* derived by the algorithm is a minimum balancing set for a given s-graph.

Fig. 5.3.12 illusrates the above algorithm applied to the graph S in Fig. 5.3.10 (Steps 1 and 2 are already shown in Fig. 5.3.10). The minimum balancing process corresponding to case (i a) of step 6 in Fig. 5.3.12 is shown in Fig.5.3.13.

Step 5 is reduced to the so-called shortest route (path) problem as follows. Let U_i be a subset of V_u whose number of elements is even. The number p of such subsets (including Φ and V_u) is equal to $2^{\#V_u-1} = 2^{2m-1}$. We make a directed graph N as follows. The vertices of N are U_1, U_2, $\cdots$, and U_p. If the conditions

$$U_j \supset U_i \text{ and } \#U_j \quad \#U_i \quad 2 \tag{5.3.91}$$

hold, then draw a directed line from the vertex U_i to U_j for i, j 1,2, ··· , p. We also attach to the line from U_i to U_j the distance between the vertices v_r and v_s, according to the calculation of step 3. The vertices v_r and v_s are defined by $\{v_r, v_s\} = U_j \quad U_i$. The vertex U_i means that the points contained in U_i have already been paired and the process from U_i to U_j means that the vertices v_r and v_s are newly paired. For the directed graph N,we find the shortest routes from the starting vertex "Φ" to the final vertex "V_u". Each of the obtained routes corresponds to an optimum pairing V_1^*, V_2^*, ···, and V_m^* given by step 5. To reduce the number of vertices in N, we restrict the routes from Φ to V_u as follows. First, v_1 is paired with another vertex v_1'. Second, if $v_1' \neq v_2$, then v_2 is paired, otherwise v_3 is paired, etc. That is, the vertex with the smallest subscript is paired at each step. This yields the condition for N such that the vertices of size 2k necessarily contain $\{v_1, v_2, \cdots, v_k\}$ for k = 1, 2, ···, m. The number of the vertices of size 2k is $\binom{2m}{k}$. Thus the total number of the vertices is $\sum_{k=0}^{m} \binom{2m - k}{k} = F_{2m}$, where F_{2m} is a Fibonacci number which is given, by Binet's formula, as follows:

$$F_{2m} = \frac{1}{\sqrt{5}}\left[\left(\frac{1 + \sqrt{5}}{2}\right)^{2m+1} - \left(\frac{1 - \sqrt{5}}{2}\right)^{2m+1}\right]. \qquad (5.3.92)$$

Hence, the number of calculations needed for seeking the optimum routes has the order of $\left(\frac{1 + \sqrt{5}}{2}\right)^{4m} \simeq 6.854^m$. In the above example, the corresponding graph N is shown by Fig. 5.3.14. In the graph, case (i) of step 5 in Fig. 5.3.12 corresponds to the shortest route $\Phi \xrightarrow{1} \{v_1, v_3\} \xrightarrow{2} \{v_1, v_2, v_3, v_5\} \xrightarrow{2} V_u$.

5.3.4 Considerations on Directions and Relative Importance of Relationships among Members

In the previous sections, we treated only the social groups in which the *directions* of member-member relationships can be disregarded or, in other words, all the relations are reciprocally symmetric (cf. Newcomb[1961] and Davis [1968])

When the directions of relations must be taken into account, social groups are represented by *signed directed graphs(signed digraphs)*. The definition of balance in this case, is given as all the *semicycles* having positive signs. A semicycle is a sequence of directed lines which constitute a circuit when the directions of the lines are disregarded and its sign is defined in the same way as the non-directed case (for details, refer to Harary, Norman, and Cartwright [1965]). Hence, there is no essential difference between directed and

non-directed cases, and all the theorems and lemmas remain valid except for some slight modifications in Sections 5.3.2 and 5.3.3. Therefore, our treatment of balancing problems is also effective by disregarding the directions of interpersonal relations (cf. Katai & Iwai[1978c]).

The modifications in Section 5.3.2 are as follows : As similar to the quantification method IV, we introduce the next symmetrized matrix.

$$B'(S) \triangleq A(S) + {}^{t}A(S) \tag{5.3.93}$$

Then, we have

$${}^{t}\mathbf{s}B'(S)\mathbf{s} \quad 2{}^{t}\mathbf{s}A(S)\mathbf{s} \tag{5.3.94}$$

for an arbitrary sign vector $\mathbf{s}$.
Hence, theorem 5.3.3 remains valid by substituting B'(S) for A(S) in the definition (5.3.20) of $\hat{d}(\mathbf{s})$. Also, theorem 5.3.4 remains valid, where λ^* is the maximum eigenvalue of B'(S). In this case, if two members (vertices) v_i and v_j have *antisymmetric* relations, i.e., $a_{ij} = -a_{ji}$, then $b_{ij}' = 0$. Hence, if there exists a member v_i in S such that all the relations between v_i and other members are antisymmetric, then $b_{ij}' = b_{ji}' = 0$ for $j = 1, 2, \ldots, n$. In such cases, we treat the problem by removing v_i from S; i.e., the assignment of v_i can be set arbitrarily for considering minimum balancing sets. For example, in the case of Fig. 5.3.15(i), v_7 can be removed from S.

The iterative algorithm in Section 5.3.2.2 is also valid by replacing A(S) with the symmetric matrix B'(S).

The only modification in Section 5.3.3 is the definition (construction) of the dual graph as follows : In a planar signed digraph S, if there exist two lines joining the same pair of vertices, then we regard the region encircled by the two lines as a face of S. For example, in the case of Fig. 5.3.15(i), the corresponding dual graph is given by Fig. 5.3.15(ii).

Also, in the previous sections, we treated the groups in which the relations between units or objects are specified merely as positive or negative. Such approaches (qualitative approaches) suit only a narrow area of actual social problems. In actual groups, the relations between units or objects must be described by the quantities reflecting their relative importance. For example, in the 3 - nation 2 affair problem in Fig. 5.2.2, the following statements are probably in accordance with common sense, at the present time; the affair *AF* (the intervention of *USSR* on African countries) is more important than the affair *C.N.* (the Chinese nuclear armaments) for the *USA*, and *USSR* is more threatened by *C.N.* than *USA* is.

In general, the relative importance of a relation corresponds to the resistance to sign change of the relation. The resistance is measured by the value of the situation due to the change. For instance, the resistance of the relation "*USSR* $\overset{(-)}{---}$ *C.N.*" is measured by the difference between the evaluation by *USSR* of the policy of opposing *C.N.* and that of the policy of approving *C.N.* Thus, the measurement of the importance is reduced to the quantification of the value of situations. In some cases, the above value is directly measured pecuniarily. However, in general, the quantification of value of situations is objectively formalized as a preference problem. For instance, the *subjectively expected utility theory* by Davidsen et al. (1957) shares a way for the quantification. For the measurement of value of situations held by a human group, it is done by psychometric methods. Among these, the method of Thurstone - Jones (1957) based on *paired comparison experiments* has high applicability to the problems in this context. For, by the method, the additivity of value assignment holds, that is, the value assignment of a composite situation is the sum of the value assignment of individual situations.

The notion of social balance is in nature qualitatively defined and hence, the characterizations of balancing sets such as lemma 5.3.1 and theorem 5.3.2 are available in this case.

The characterization of the minimum balancing sets and the algorithm to derive the sets for this case have no essential difference with those described previously but slight modifications must be taken into account as follows (cf. Katai & Iwai[1978a]) :

By assumming the additivity of the values attached to the relations, theorem 5.3.3 for the characterization of the minimum balancing set is modified by transforming the adjacency matrix $A(S) = (a_{ij})$ into

$$A(S) = (a_{ij} \cdot w_{ij}), \tag{5.3.95}$$

where a_{ij} is the sign and w_{ij} is the relative importance of the relation $v_i v_j$. The other descriptions in the theorem remain valid.

The iterative algorithm in Section 5.3.2.2 remains valid by the above modification of A(S).

For example, we show an application of the above method to the problem of the so-called coalition formation of the units. In this case, the problems are treated so that the unit-unit relations are fixed as positive (+) and the signs of the unit-object relations are changed to make the groups balanced. This means that extremely high relative importance is given to the unit-unit

relations. Let us consider a group composed of three units v_1, v_2 and v_3 (such as *USA*, *USSR* and *China*) and two objects v_4 and v_5 (such as *C.N.* and *AF*) as illustrated in Fig. 5.3.16. The candidates of the optimum sign vector t_{s*} in theorem 5.3.3 are : *CASE* 1; (1 1 1 1 1) (the objects v_4 and v_5 are adopted by the three units), *CASE* 2; (1 1 1 1 -1) (v_4 is adopted and v_5 is rejected), *CASE* 3; (1 1 1 -1 1) (v_4 is rejected and v_5 is adopted) and *CASE* 4; (1 1 1 -1 -1) (both v_4 and v_5 are rejected). According to the theorem, we introduce the following indices :

$$\left.\begin{array}{ll} w(v_4) = w(v_4\text{-}v_1) + w(v_4\text{-}v_2) & w(v_4\text{-}v_3) \\ w(v_5) = w(v_5\text{-}v_1) + w(v_5\text{-}v_3) & w(v_5\text{-}v_2) \end{array}\right\} . \qquad (5.3.96)$$

Which of the above four cases corresponds to the minimum balancing set is illustrated in Fig. 5.3.17. The object-object relation $(v_4\text{-}v_5)$ produces the contraction or the expansion of the regions corresponding to the four cases. For example, the contraction of the region of *CASE* 2 in Fig. 5.3.17(b) is due to the supplemented resistance to change sign of the object-object relation from positive to negative. The neighbourhood of the origin : $w(v_4) \cong 0$ and $w(v_5) \cong 0$ means that the rejecting power of the object v_4 $(w(v_4\text{-}v_i) + w(v_4\text{-}v_2))$ and adopting power of $v_4(w(v_4\text{-}v_3))$ cancel each other and those of v_5 also cancel each other. Therefore, in the above situation, there is no conflict for the change of the balancing modes among cases 1, 2, 3 and 4 in Fig. 5.3.17(a) (without an object-object relation). On the other hand, in Fig. 5.3.17(b) or (c) (with an object-object relation), only cases 1 and 4 (for (b)) or cases 2 and 3 (for (c)) are possible.

The modification in Section 5.3.3 is the alteration of the dual graph. The definition of the dual graph is altered such that, in the new dual graph, each line has the value (length) which is equal to the relative importance of the corresponding relation in the original plane graph. For example, the line v_1v_6 in Fig. 5.3.11(b) has the value $\bar{w}_{16}$ which is the same as the relative importance w(*USA C.N.*). The length of a path in the dual graphs is defined as the sum of the values of the lines contained in it. Therefore, the optimum path L_1^*, L_2^*, ..: , L_m^* defined by (5.3.62') and the optimum pairing V_1^*, V_2^*, ... , V_m^* defined by (5.3.79) may be different from the case of section 5.3.3 and determined by the relative importances of the relations. For example, for the 3 - nation 2 - affair group in Fig. 5.2.2(a), the corresponding dual graph is shown in Fig. 5.3.18,where $\bar{w}_{ij}$ is the value of the corresponding relation in the original group. The following are the six paths

joining v_1 and v_2. (1); $v_1 - v_4 - v_2$ (the concession of *USA* to *USSR*), (2); $v_1 - v_4 - v_5 - v_2$ (the concession of *USSR* to *USA*), (3); $v_1 - v_4 - v_5 - v_3 - v_6 - v_2$ (the concession of *USSR* and *China* to *USA*), (4); $v_1 - v_3 - v_5 - v_2$(the concession of *USSR* and *China* each other), (5); $v_1 - v_3 - v_6 - v_2$(the concession of *China* to *USA*) and (6); $v_1 - v_6 - v_2$ (the concession of *USA* to *China*). When we have the following relative importances : *C.N.-USA* is less than *C.N.-USSR* (as previously stated), *USSR-C* is larger than *USA-C*, and *C-C.N.* is larger than *C-AF*; then it follows that $\bar{w}_{16} < \bar{w}_{13}$, $\bar{w}_{35} > \bar{w}_{26}$, $\bar{w}_{36} > \bar{w}_{25}$. From Fig. 5.3.18, it is clear that paths (3), (4) and (5) cannot be the minimum balancing sets. The remaining candidates are (1), (2) and (6), while in the discussion in Section 5.3.3, the minimum balancing sets are (5) and (6). If the importance of the relations between *USA* and *USSR* ($\bar{w}_{14} + \bar{w}_{42}$ and $\bar{w}_{14} + \bar{w}_{45} + \bar{w}_{52}$) is higher than the importance of the relation of *USA* to *China* ($\bar{w}_{16} + \bar{w}_{62}$), then the minimum balancing set is uniquely determined as (6). In this case, *USA* concedes to *China* and *USSR* is isolated. In the other case, (1) or (2) is the minimum balancing set according to whether the item *AF* is less important for *USA* than for *USSR* and *China* or not, that is,

$$\bar{w}_{24} \lessgtr \bar{w}_{45} + \bar{w}_{52} , \tag{5.3.97}$$

where the upper inequality corresponds to (1) (*USA* concedes to *USSR*) and the lower to (2) (*USSR* concedes to *USA*).

5.3.5 Derivation of the Minimal Balancing sets of Unbalanced Group Structures

As mentioned previously, there exists another kind of balancing set introduced by Harary, called minimal balancing set. A minimal balancing L_b^m for an s-graph S is defined as a balancing set of S whose proper subset is no longer a balancing set for S. The notion of block decomposition in graph theory is also beneficial in searching for the minimal balancing sets for an s-graph. As similar to theorem 5.3.5, the following theorem holds.

THEOREM 5.3.8 : Let S be an s-graph and B_1, B_2, $\cdots$, and B_t be its block components. Then a set L of lines of S is a minimal balancing set of S iff L is represented as

$$L \quad \bigcup_{i=1}^{t} L_i , \tag{5.3.98}$$

where L_i is a minimal balancing set for $i = 1, 2, \cdots, t$.

Therefore, we can assume that given s-graphs are nonseparable, and hence connected s-graphs. First of all, we consider the characterization problem of the minimal balancing sets for arbitrarily given s-graphs. It should be noted that the sign vectors $\mathbf{s}$ and $-\mathbf{s}$ ($(-s_1, -s_2, \cdots, -s_n)$) yield the same grouping of V. Hence, we identify $\mathbf{s}$ with $-\mathbf{s}$. By the identification, the following lemma holds.

LEMMA 5.3.8 : If S is connected, then each sign vector corresponds to a balancing set for an s-graph S in a one-to-one manner, i.e., the following holds.

$$\text{if } a_{ij}s_is_j = a_{ij}s_i's_j' \text{ for all } i, j \quad 1,2, \cdots, n,$$

then

$$\mathbf{s}' \quad \mathbf{s} \text{ (i.e., } \mathbf{s}' = \mathbf{s} \text{ or } \mathbf{s}' = \quad \mathbf{s}). \tag{5.3.99}$$

PROOF : Suppose that $\mathbf{s}' \neq \mathbf{s}$. Then there exist i and j such that

$$s_i' \quad s_i \text{ and } s_j' = \quad s_j'. \tag{5.3.100}$$

From the condition that S is connected, there exists a path $v_iv_{q_1}$, $v_{q_1}v_{q_2}$, $\cdots$, $v_{q_k}v_j$ joining v_i and v_j. Let us calculate the following quantities:

$$(a_{iq_1}s_is_{q_1})\cdot(a_{q_1q_2}s_{q_1}s_{q_2})\cdot \cdots \cdot(a_{q_kj}s_{q_k}s_j) = a_{iq_1}a_{q_1q_2}\cdots a_{q_kj}s_is_j, \tag{5.3.101}$$

$$(a_{iq_1}s_i's_{q_1}')\cdot(a_{q_1q_2}s_{q_1}'s_{q_2}')\cdot \cdots \cdot(a_{q_kj}s_{q_k}'s_j')$$

$$= a_{iq_1}a_{q_1q_2}\cdots a_{q_kj}s_i's_j'. \tag{5.3.102}$$

From (5.3.100), these quantities have different values. Therefore, there exist q_t and q_{t+1}, such that

$$a_{q_tq_{t+1}}s_{q_t}s_{q_{t+1}} \neq a_{q_tq_{t+1}}s_{q_t}'s_{q_{t+1}}'. \tag{5.3.103}$$ □

To characterize the minimal balancing sets for S by sign vectors, we introduce a binary relation on the set of n-dimensional sign vectors. The relation is defined in connection with adjacency matrix A(S), hence we denote it by "$\underset{A(S)}{\gtrdot}$".

DEFINITION 5.3.2 : Let $\mathbf{s}$ and $\mathbf{s}'$ be two arbitrary n-dimensional sign vectors. Then $\mathbf{s} \underset{A(S)}{\gtrdot} \mathbf{s}'$, iff the following condition is satisfied:

$$a_{ij}s_is_j \geq a_{ij}s_i's_j' \quad \text{for all } i, j = 1, 2, \cdots, n. \tag{5.3.104}$$

From lemma 5.3.8, the binary relation has the following property.

THEOREM 5.3.9 : If S is connected , then $\overset{\succcurlyeq}{A(S)}$ is a partial ordering, i.e., it is reflexive, transitive, and antisymmetric.

PROOF : It is obvious from the definition that the relation is reflexive and transitive. Suppose that $\mathbf{s} \overset{\succcurlyeq}{A(S)} \mathbf{s}'$ and $\mathbf{s}' \overset{\succcurlyeq}{A(S)} \mathbf{s}$, then $a_{ij}s_i's_j' = a_{ij}s_is_j$ for all i, j = 1, 2, ···, n. From lemma 5.3.8, we have $\mathbf{s}' = \mathbf{s}$. Hence $\overset{\succcurlyeq}{A(S)}$ is antisymmetric. □

The partial ordering characterizes the minimal balancing sets as follows:

LEMMA 5.3.9 : L_b is a minimal balancing set of S, iff the corresponding sign vector $\mathbf{s}$ is maximal in the sense of the partial ordering "$\overset{\succcurlyeq}{A(S)}$", i.e., there exists no other $\mathbf{s}'$ such that $\mathbf{s}' \overset{\succcurlyeq}{A(S)} \mathbf{s}$.

PROOF : Let $\mathbf{s}'$ be an arbitrary sign vector and L_b' be the corresponding balancing set. Then, from theorem 5.3.2, we have $L_b' \subset L_b$, iff $a_{ij}s_is_j = -1$ for any i and any j such that $a_{ij}s_i's_j' = -1$. That is $L_b' \subset L_b$ iff $\mathbf{s}' \overset{\succcurlyeq}{A(S)} \mathbf{s}$. □

For a balancing set L_b of an s-graph S, we introduce another binary relation (an equivalence relation) "$\overset{\sim}{L_b}$" on the set V of the vertices of S as follows.

DEFINITION 5.3.3 : Let v_i and v_j be vertices of an s-graph S and L_b be a balancing set for S. Then we define

$$v_i \overset{\sim}{L_b} v_j \quad \text{iff } v_i = v_j \text{ or there exists a path joining } v_i \text{ and } v_j \text{ such that any line in the path is not contained in } L_b. \tag{5.3.105}$$

By the above equivalence relation, set V is partitioned into equivalence classes which are disjoint with each other. Let $E(v_i)$ denote the equivalence class containing v_i.

We classify the lines of S into two categories in relation to the equivalence classes corresponding to a balancing set L_b.

DEFINITION 5.3.4 : We call a line v_iv_j in L_b an *inherent* line, iff both v_i and v_j are contained in the same equivalence class induced by L_b. We call a

line $v_i v_j$ a *non-inherent* line, iff v_i and v_j belong to different equivalence classes.

From lemma 5.3.9, the characterization of the minimal balancing set of S is given by the next theorem (cf. Katai & Iwai[1978c])

THEOREM 5.3.10 : If S is connected, then the following statements are equivalent.

1) L_b is a minimal balancing set.
2) Every line in L_b is inherent.
3) L_b yields only one equivalence class, i.e., the set V.

PROOF : 1) → 2) :

Suppose that L_b contains a non-inherent line $\{v_p, v_q\}$. Then, we have

$$E(v_p) \cap E(v_q) = \Phi. \tag{5.3.106}$$

Let $\mathbf{s}'$ be a sign vector as follows:

$$\begin{aligned} s_i' &= -s_i \quad \text{if } v_i \in E(v_p), \\ s_i' &= s_i \quad \text{if } v_i \notin E(v_p). \end{aligned} \tag{5.3.107}$$

From (5.3,106),

$$\mathbf{s}' \neq \mathbf{s} \text{ (i.e., } \mathbf{s}' \neq \mathbf{s} \text{ and } \mathbf{s}' \neq -\mathbf{s}). \tag{5.3.108}$$

It is obvious that

$$\begin{aligned} a_{ij}s_i's_j' &= -a_{ij}s_is_j \quad \text{if either } v_i \text{ or } v_j \text{ is contained in } E(v_p) \text{ and} \\ &\qquad \text{the other is not contained in } E(v_p), \text{ and} \\ a_{ij}s_i's_j' &= a_{ij}s_is_j \quad \text{if otherwise.} \end{aligned} \tag{5.3.109}$$

If v_i and v_j belong to different equivalence classes, then $a_{ij}s_is_j = 0$ or -1. Hence, $a_{ij}s_i's_j' \geqq a_{ij}s_is_j$ for all i and j = 1, 2, ··· , n, i.e.,

$$\mathbf{s}' \underset{A(S)}{\geqslant} \mathbf{s}. \tag{5.3.110}$$

From (5.3.108) and (5.3.110), $\mathbf{s}$ is not maximal. That is, if $\mathbf{s}$ is maximal, then L_b does not contain non-inherent lines.

2) → 3):

Suppose that L_b yields more than one equivalence class, say $E_1, E_2, \cdots, E_r$. Then, from the supposition that S is connected, E_i and E_j $(i \neq j)$ are joined by a line in L_b for some i and j. The line is obviously non-inherent.

3) → 1):

Suppose that L_b is not minimal. Let $\mathbf{s}$ be the corresponding sign vector. Then, from lemma 5.3.9, there exists a sign vector $\mathbf{s}'$ such that

$$\mathbf{s}' \overset{\gg}{A(S)} \mathbf{s} \text{ and } \mathbf{s}' \neq \mathbf{s} . \tag{5.3.111}$$

Let V' and V'' be as follows:

$$\begin{aligned} V' &\quad \{v_i \mid v_i \in V,\ s_i s_i' \quad 1\} \neq \Phi , \\ V'' &\quad \{v_i \mid v_i \in V,\ s_i s_i' \quad -1\} \neq \Phi . \end{aligned} \tag{5.3.112}$$

Suppose that there exists a path joining a vertex in V' and a vertex in V'' such that each line in the path is not contained in L_b. Let the path be denoted by $v_t v_{q_1}$, $v_{q_1} v_{q_2}, \cdots, v_{q_k} v_u$. Let us calculate the following quantities:

$$\begin{aligned} &(a_{tq_1} s_t s_{q_1}) \cdot (a_{q_1 q_2} s_{q_1} s_{q_2}) \cdot \cdots \cdot (a_{q_k u} s_{q_k} s_u) \\ &\qquad = a_{tq_1} a_{q_1 q_2} \cdots a_{q_k u} s_t s_u , \end{aligned} \tag{5.3.113}$$

$$\begin{aligned} &(a_{tq_1} s_t' s_{q_1}') \cdot (a_{q_1 q_2} s_{q_1}' s_{q_2}') \cdot \cdots \cdot (a_{q_k u} s_{q_k}' s_u') \\ &\qquad a_{tq_1} a_{q_1 q_2} \cdots a_{q_k t} s_t' s_u' . \end{aligned} \tag{5.3.114}$$

From the condition on the path,

$$a_{tq_1} s_t s_{q_1} \quad a_{q_1 q_2} s_{q_1} s_{q_2} = \cdots = a_{q_k u} s_{q_k} s_u = 1 . \tag{5.3.115}$$

Also, from (5.3.111), we have

$$a_{tq_1} s_t' s_{q_1}' = a_{q_1 q_2} s_{q_1}' s_{q_2}' = \cdots = a_{q_k u} s_{q_k}' s_u' \quad 1 . \tag{5.3.116}$$

That is, the left-hand sides of (5.3.113) and (5.3.114) have the same value. On the other hand, from (5.3.112), the right-hand sides of (5.3.113) and (5.3.114) do not coincide. Hence, for an arbitrary pair of vertices $v_i \in V'$ and $v_j \in V'$, $v_i \not\sim_{L_b} v_j$. That is, if L_b yields only one equivalence class, then L_b is minimal. □

The meaning of " inherent " and " non-inherent " is given by the next theorem.

THEOREM 5.3.11 : Let L_b be a balancing set for S and L_b^i be the set of inherent lines in L_b. Then, the following statements hold.

If L_b' is a subset of L_b and is also a balancing set for S, then

$$(L_b \supset)\ L_b' \supset L_b^i \ . \tag{5.3.117}$$

Particularly, any minimal balancing set L_b^m contained in L_b also contains L_b^i.

If x is a non-inherent line of L_b, then there exists a balancing set L_b' such that

$$L_b - x \supset L_b' \ . \tag{5.3.118}$$

PROOF : Suppose that $L_b' \subset L_b$. Let $\mathbf{s}'$ be the corresponding sign vector of L_b'. Then, we have $\mathbf{s}' \underset{A(S)}{\gtrdot} \mathbf{s}$. Let V' and V'' be the sets defined by (5.3.112). Then, from the proof of theorem 5.3.10 (the proof of 3) → 1)), $E(v_i) \subset$ V' or $E(v_i) \subset$ V'' for an arbitrary vertex v_i of S. That is, the following statement holds.

$$\text{If } L_b \supset L_b', \text{ i.e., } \mathbf{s}' \underset{A(S)}{\gtrdot} \mathbf{s}, \text{ then } s_i' = s_i \text{ or } s_i' = -s_i$$
$$\text{for all } v_i \text{ in an equivalence class induced by } L_b. \tag{5.3.119}$$

Therefore, we obtain

$$a_{ij}s_i's_j' = a_{ij}s_is_j \text{ for all } v_i \text{ and } v_j \text{ in the same equivalence class}$$
$$\text{induced by } L_b. \tag{5.3.120}$$

Hence, if v_iv_j is an inherent line of L_b, then v_iv_j is contained in L_b'.

Suppose that v_tv_u is a non-inherent line of L_b. Then, from the proof of theorem 5.3.10 (the proof of 1) → 2)), there exists a sign vector $\mathbf{s}'$ such that

$$\mathbf{s}' \underset{A(S)}{\gtrdot} \mathbf{s} \text{ and } a_{tu}s_t's_u' = -a_{tu}s_ts_u = 1 \ . \tag{5.3.121}$$

Let L_b' be the corresponding balancing set. Then, we have $L_b' \subset L_b$ and $v_tv_u \notin L_b'$. □

In the sequel, we consider a method to derive the minimal balancing sets contained in an arbitrarily given balancing set L_b. From lemma 5.3.9, the problem is reduced to search for the maximal sign vectors $\mathbf{s}$'s such that $\mathbf{s}' \underset{A(S)}{\gtrdot} \mathbf{s}$, where $\mathbf{s}$ is the corresponding sign vector of L_b. Let $\hat{\mathbf{t}} = {}^t(\hat{t}_1, \hat{t}_2, \cdots, \hat{t}_n)$ be denoted as follows:

$$\hat{t}_i = s_i's_i^{-1}, \text{ i.e., } s_i' \quad s_i\hat{t}_i, \quad i = 1, 2, \cdots, n, \tag{5.3.122}$$

Then, from (5.3.119) in the proof of theorem 5.3.11.

$$\text{if } \mathbf{s}' \underset{A(S)}{\gtrdot} \mathbf{s}, \text{ then } \hat{t}_i \equiv 1 \text{ or } \hat{t}_i \equiv -1$$

in each equivalence class induced by L_b. (5.3.123)

If we seek the optimal $\hat{\mathbf{t}}$ (which yields maximal $\mathbf{s}'$) instead of searching for $\mathbf{s}'$, then we can condense each equivalence class into a vertex. That is, we construct a graph G_c in which each vertex e_i corresponds to an equivalence class E_i induced by L_b, and two vertices are joined by a line when there exists a non-inherent line of L_b joining the corresponding equivalence classes. Let t_j denote the corresponding assignment to e_j for $j = 1, 2, \cdots, k$, and E_{r_i} denote the equivalence class containing v_i for $i = 1, 2, \cdots, n$. Then

$$\hat{t}_i = t_{r_i}, \quad s_i' = s_i t_{r_i} \quad \text{for } i = 1, 2, \cdots, n. \tag{5.3.124}$$

Fig. 5.3.19 shows the above construction of graph G_c for the s-graph S in Fig. 5.3.4. A balancing set L_b and the corresponding sign vector are shown in (i). In this case, L_b yields four equivalence classes E_1, E_2, E_3, and E_4. Lines v_2v_3 and v_6v_7 are contained in E_1 and E_3, respectively, hence they are inherent lines of L_b, and the other broken lines are non-inherent lines of L_b. The final graph G_c is shown in (ii), where e_i corresponds to equivalence class E_i for $i = 1, 2, 3$, and 4.

LEMMA 5.3.10 : An assignment vector $\mathbf{t} = {}^t(t_1, t_2, \cdots, t_k)$ gives a minimal balancing set L_b' contained in L_b, iff subgraph G_c' of G_c is a connected graph, where G_c' is obtained by removing all the lines e_pe_q from G_c such that $t_p \neq t_q$.

PROOF : From (5.3.124),

$$\begin{aligned} &s_i's_j' = s_is_j \quad \text{if } t_{r_i} = t_{r_j}, \\ &s_i's_j' \neq s_is_j \quad \text{if } t_{r_i} \neq t_{r_j}. \end{aligned} \tag{5.3.125}$$

Therefore, a non-inherent line of L_b joining equivalence classes with the same assignment is also contained in L_b', and a non-inherent line joining equivalence classes with different assignments is not contained in L_b'. Hence, two equivalence classes (induced by L_b) which have different assignments and are adjacent to each other (i.e., are joined by a non-inherent line of L_b) belong to the same equivalence class of L_b'. The lemma follows from theorem 5.3.10.□

The following theorem is a direct consequence of the above lemma.

THEOREM 5.3.2 : An assignment vector $\mathbf{t} = {}^t(t_1, t_2, \cdots, t_k)$ gives a mini-

mal balancing set L_b' contained in L_b, iff there exists a spanning tree of G_c such that $t_p = t_q$ when the distance between e_p and e_q on the tree is even and $t_p = -t_q$ when the distance is odd.

PROOF : From lemma 5.3.10, $\mathbf{t}$ gives a minimal balancing set, iff G_c' is a connected spanning subgraph of G_c.

It is obvious from the definition of G_c', if an assignment vector $\mathbf{t} = {}^t(t_1, t_2, \cdots, t_k)$ is given by a spanning tree of G_c such that $t_p = t_q$ when the distance between e_p and e_q on the tree is even and $t_p = -t_q$ when the distance is odd, the corresponding graph G_c' contains the tree; i.e., G_c' is a connected spanning subgraph of G_c.

If graph G_c' is a connected spanning subgraph of G_c, then G_c' contains a spanning tree of G_c.

From the definition of G_c', two arbitrary vertices (of G_c) which are adjacent to each other in the tree have different assignments. □

From the above theorem, the problem is reduced to searching for the spanning trees of graph G_c. Each spanning tree of a graph is obtained by the following procedures (for details, refer to Harary [1969, pp. 32 - 40]). First, remove a line contained in a circuit from G_c. Second, remove a line which is contained in a circuit from the graph obtained by the previous step. Repeat the above process until all the circuits in G_c disappear.

In Fig. 5.3.19(ii), graph G_c contains two circuits x_1, x_2, x_3 and x_3, x_4, x_5. If line x_1 is removed from G_c, then the former circuit disappears. The removal of x_3 eliminates the remaining circuit x_3, x_4, x_5. No circuits remain and the procedure stops. Table 5.3.2 shows the list of the lines removed from G_c. There are eight ways for eliminating the two circuits of G_c; case 1 corresponds to the above mentioned example. Fig. 5.3.20 shows corresponding spanning trees and the assignment vectors.

In general, duplications of the assignment vectors occur. In the above example, four cases 1, 4, 7, and 8 yield the same assignment vector $\mathbf{t} = {}^t(1, -1, -1, 1)$. Hence, we obtain five distinct assignment vectors and the corresponding minimal balancing sets are shown in Fig. 5.3.21.

The algorithm to derive all the minimal balancing sets contained in an arbitrary given balancing set L_b for an s-graph S is summarized as follows (cf. Katai & Iwai[1978c]):

step 1: Partition the set V of the vertices of S into the equivalence classes induced by L_b.

step 2: Make the condensed graph G_c.

step 3: Seek all the spanning trees of G_c.

step 4: Calculate the assignment vector $\mathbf{t}$ for each sapnning tree, where a specified vertex, say e_1, is assigned as +1 (to avoid duplications).

step 5: Delete the duplicated assignment vectors, i.e., list up all the distinct assignment vectors derived from step 4.

step 6: Referring to each assignment vector $\mathbf{t}$, delete all the non-inherent lines from L_b, each of which joins two equivalence classes with different assignment values of $\mathbf{t}$.

The final step 6 yields all the minimal balancing sets contained in L_b.

For planar s-graphs, theorem 5.3.6 yields the following theorem.

THEOREM 5.3.13 : Let S and V_u be the same as in theorem 5.3.6. Then L_b is a minimal balancing set for S iff there exist sets $L_1, L_2, \cdots, L_m$ of lines of D(S) and sets (pairs) $V_1 \{v_{p_1}, v_{p_1'}\}$, $V_2 \{v_{p_2}, v_{p_2'}\}$, $\cdots$, and $V_m = \{v_{p_m}, v_{p_m'}\}$ of vertices of D(S) such that conditions (5.3.58) (5.3.62) of theorem 5.3.6 are satisfied (i.e., the set of lines L_{m+1} in theorem 5.3.6 is empty).

PROOF : The theorem is easily verified from the fact that set L_{m+1} in theorem 5.3.6 is unnecessary for balancing and any proper subset of a path is no longer a path joining the endpoints of the path. □

From lemma 5.3.7 and theorem 5.3.13, the following procedure gives a minimal balancing set. First, partition the set V_u of negative vertices into pairs $V_1, V_2, \cdots$, and V_m. Second, seek the sets $L_1, L_2, \cdots$, and L_m such that each L_i is composed of the lines contained in a path joining the vertices of V_i for i = 1, 2, $\cdots$, m. Third, calculate $\sigma_1(L_1) + \sigma_1(L_2) + \cdots + \sigma_1(L_m)$. Let L be the lines contained in the above sum. Then, $d^{-1}(L)$ is a balancing set. If L contains no cycles, then $d^{-1}(L)$ is also a minimal balancing set. If L contains at least one cycle, then delete all the lines (each of which is contained in a cycle in L) from L. The resultant set gives a minimal balancing set (see theorem 5.3.13).

Also, from theorem 5.3.13, an arbitrary minimal balancing set is given by the above procedure. For example, in the case of the s-graph shown in Fig. 5.3.10(i), let $V_1 = \{v_1, v_4\}$, $V_2 = \{v_2, v_3\}$, $V_3 = \{v_5, v_6\}$, and L_1 $\{v_1v_3, v_3v_{12}, v_{12}v_4\}$, $L_2 = \{v_2v_{10}, v_{10}v_{11}, v_{11}v_{12}, v_{12}v_4, v_4v_8, v_8v_3\}$, $L_3 = \{v_5v_{11}, v_{11}v_{12}, v_{12}v_4, v_4v_{13}, v_{13}v_6\}$. Then, we have

$$\sigma_1(L_1) + \sigma_1(L_2) + \sigma_1(L_3) = v_1v_3 + (v_2v_{10} + v_{10}v_{11} + v_{11}v_5) + (v_4v_{13} + v_{13}v_6) + (v_3v_{12} + v_{12}v_4 + v_4v_8 + v_8v_3). \quad (5.3.126)$$

The corresponding balancing set is shown in Fig. 5.3.22. In the above equation, the right-hand side is composed of three paths joining v_1 and v_3, v_2 and v_5, v_4 and v_6 and a cycle v_3v_{12}, $v_{12}v_4$, v_4v_8, v_8v_3. Deleting the cycle from the right hand side yields a minimal balancing set as in Fig. 5.3.23. Continuing the above procedure, we can find all the balancing and the minimal balancing sets.

5.4 Quantification of Degree of Balance from Statistical Point of View and Classification of Types of Balance by Finite-state Systems Theoretical Aspect

In previous sections, we have discussed the derivation methods of the minimum balancing sets for unbalanced group structures. As mentioned in Section 5.2, the number of elements in a minimum balancing set of an unbalanced group can be regarded as a degree of balance (more precisely, degree of unbalance) of the group. Thus, the derivation methods provide a way for calculating the degree of balance.

These discussions are, however, concerned only with the characterization of given fixed social groups (s-graphs). In actual social groups, the member-member relations are not constant but fluctuate between positive and negative. Therefore, the degree of social balance should be evaluated statistically and also a finite-state systems theoretic structure of the interpersonal relations should be introduced for the definition of social balance (cf. Katai, Iwai, et al.[1975a&b,1976d] and Katai & Iwai[1978b]).

In section 5.4.1, we calculate the probability of balance of social groups with given graph structures assuming that the member-member relations are stochastically independent. For the calculation, we introduce the notion of *cycle polynomial* and the statistical properties of social balance are discussed.

The cases where the relations are not independent are discussed in Section 5.4.2. We introduce a group structure called *relational structure* where the relation between a pair of members is determined by the opinions or attitudes they hold about certain social arguments, beliefes or norms. In this framework, three types of social balance are introduced, i.e., *totally balanced*,

totally unbalanced and *partially balanced*. The first type is the case where there is no room for the occurrence of an unbalanced state. The second type is one where there is no possibility of resolution of an unbalanced state. A group of the third type may be balanced or unbalanced according to the opinions held by its members. We also consider the characterization of three types.

5.4.1 Quantification of Degree of Balance from Statistical Point of View

In this section, the statistical nature of the graph structures of social groups on their balance is discussed. We denote the graph (group structure) of a social group by G and the lines in G by $x_1, x_2 \cdots, x_{m-1}$ and x_m. Also the cycles in G are denoted by $\sigma_0, \sigma_1, \cdots, \sigma_{k-1}$, where a cycle of G is a sequence of lines composed of closed paths (circuits) contained in G. The cycle composed of no circuits is called 0-cycle. Therefore, a circuit is a cycle but not *vice versa*. For example, in Fig. 5.4.1, the sequence of lines $x_1, x_2, \ldots, x_8$ is not a circuit but a cycle composed of two circuits, i.e., x_1, x_2, x_3, x_4 and x_5, x_6, x_7, x_8. For the 0-cycle σ_0, we set $\sigma_0 \quad 1$. A cycle σ_i in G is formally described by the product of lines, that is,

$$\sigma_i = x_{i_1} \cdot x_{i_2} \cdot \ \cdots \ \cdot x_{i_{m_i}}, \tag{5.4.1}$$

where distinct lines $x_{i_1}, x_{i_2}, \ldots$ and $x_{i_{m_i}}$ are contained in σ_i. Using the above rule, we introduce the following definition (cf. Katai, Iwai, et al[1975b & 1976d]).

DEFINITION 5.4.1 : For an arbitrary graph G, the following polynomial of x_1, $x_2, \ldots$ and x_m is said to be the *cycle polynomial* of G.

$$c(x_1, x_2, \ldots, x_m : G) \quad \sum_{i=0}^{k-1} \sigma_i = 1 + \sum_{i=1}^{k-1} x_{i_1} \cdot x_{i_2} \cdot \ \cdots \ \cdot x_{i_{m_i}}. \tag{5.4.2}$$

The corresponding cycle polynomial of graph G in Fig. 5.4.1 is calculated as follows:

$$\begin{aligned} & c(x_1, x_2, \ldots, x_m : G) \\ & = 1 + x_1 \cdot x_2 \cdot x_3 \cdot x_4 + x_5 \cdot x_8 \cdot x_9 + x_6 \cdot x_7 \cdot x_9 + x_{10} \cdot x_{11} \cdot x_{12} + x_5 \cdot x_6 \cdot x_7 \cdot x_8 \\ & \quad + x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_8 \cdot x_9 + x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_6 \cdot x_7 \cdot x_9 \\ & \quad + x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_{10} \cdot x_{11} \cdot x_{12} + x_5 \cdot x_8 \cdot x_9 \cdot x_{10} \cdot x_{11} \cdot x_{12} \end{aligned}$$

$$+ x_6 \cdot x_7 \cdot x_9 \cdot x_{10} \cdot x_{11} \cdot x_{12} + x_5 \cdot x_6 \cdot x_7 \cdot x_8 \cdot x_{10} \cdot x_{11} \cdot x_{12}$$
$$+ x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_6 \cdot x_7 \cdot x_9 \cdot x_{10} \cdot x_{11} \cdot x_{12}$$
$$+ x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_8 \cdot x_9 \cdot x_{10} \cdot x_{11} \cdot x_{12} + x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6 \cdot x_7 \cdot x_8$$
$$+ x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6 \cdot x_7 \cdot x_8 \cdot x_{10} \cdot x_{11} \cdot x_{12}. \qquad (5.4.3)$$

As aforementioned, a collection of cycles $\{\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_u}\}$ is called a set of fundamental cycles (cycle basis) of G iff any cycle σ of G is uniquely represented by a product of cycles contained in the collection under the following rule :

$$x_i^2 = 1 \qquad \text{for all } i = 1,2,\ldots,m. \qquad (5.4.4)$$

The above rule corresponds to the rule : 1 + 1 = 0 on the two element field $F_2 = \{0, 1\}$. In Fig. 5.4.1, the collection of cycles $\{x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6 \cdot x_7 \cdot x_8$, $x_5 \cdot x_6 \cdot x_7 \cdot x_8$, $x_5 \cdot x_8 \cdot x_9$, $x_5 \cdot x_6 \cdot x_7 \cdot x_8 \cdot x_{10} \cdot x_{11} \cdot x_{12}\}$is a set of fundamental cycles. For example, a cycle $x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_6 \cdot x_7 \cdot x_9 \cdot x_{10} \cdot x_{11} \cdot x_{12}$ is represented by the product of all the cycles contained in the set.

From the above definition, any cycle of G is represented by a product of circuits of G. Therefore, we can assume that all the cycles $\sigma_{i_1}, \sigma_{i_2}, \ldots$ and σ_{i_u} are circuits. For example, $\{x_1 \cdot x_2 \cdot x_3 \cdot x_4$, $x_5 \cdot x_8 \cdot x_9$, $x_6 \cdot x_7 \cdot x_9$, $x_{10} \cdot x_{11} \cdot x_{12}\}$ constitues a set of fundamental cycles of G in Fig. 5.4.1.

Using the notion of fundamental cycles, the calculation of a cycle polynomial is done as follows:

LEMMA 5.4.1 : Let $\{\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_u}\}$ be a set of fundamental cycles of G. Then the following equality holds.

$$c(x_1, x_2, \ldots, x_m : G) = \prod_{j=1}^{u} (1 + \sigma_{i_j}). \qquad (5.4.5)$$

Therefore, the number of cycles k is given by

$$k \quad 2^u. \qquad (5.4.6)$$

PROOF : Each cycle in G is uniquely represented by a product of cycles contained in the set $\{\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_u}\}$. Also, from the definition of cycle, any product of cycles in G is also a cycle of G. Therefore, equalities (5.4.5) and (5.4.6) are valid. □

In the case of the above example, we have

$$c(x_1,x_2,\ldots,x_{12}) = (1 + x_1 \cdot x_2 \cdot x_3 \cdot x_4)(1 + x_5 \cdot x_8 \cdot x_9)(1 + x_6 \cdot x_7 \cdot x_9)$$
$$(1 + x_{10} \cdot x_{11} \cdot x_{12}). \qquad (5.4.7)$$

A cycle polynomial has the following property.

LEMMA 5.4.2 : Let σ be a cycle of G. Then under the rule (5.4.4), the following equality holds.

$$\sigma \cdot c(x_1,x_2,\ldots,x_m : G) = c(x_1,x_2,\ldots,x_m : G). \qquad (5.4.8)$$

PROOF : Let σ and σ' be cycles of G. Then under the rule (5.4.4), their product $\sigma \cdot \sigma'$ is also a cycle of G. If two cycles $\sigma \cdot \sigma'$ and $\sigma \cdot \sigma''$ coincide with each other, then σ' is equal to σ'', for

$$\sigma'' = (\sigma \cdot \sigma) \cdot \sigma'' = \sigma \cdot (\sigma \cdot \sigma'') = \sigma \cdot (\sigma \cdot \sigma') \quad (\sigma \cdot \sigma) \cdot \sigma' - \sigma'. \qquad (5.4.9)$$

Therefore, the polynomial $\sigma \cdot c(x_1,x_2,\ldots,x_m : G)$ is the sum of distinct k cycles of G. The number of cycles of G is equal to k. Therefore, $\sigma \cdot c(x_1,x_2,\ldots,x_m : G)$ $c(x_1,x_2,\ldots,x_m : G)$. □

The following theorem is a derect consequence of the above lemma.

THEOREM 5.4.1 : When a value d_i (+1 or -1) is attached to the line x_i for i 1,2,...,m, then the following statements hold.

$$G \text{ is balanced iff } c(d_1,d_2,\ldots,d_m : G) = k, \qquad (5.4.10)$$

$$G \text{ is unbalanced iff } c(d_1,d_2,\ldots,d_m : G) = 0. \qquad (5.4.11)$$

PROOF : G is balanced iff each circuit σ has a positive sign. According to the definition of cycle, the sign of each cycle in G is also positive. Therefore, from the definition of cycle polynomial, $c(d_1,d_2,\ldots,d_m : G)$ is equal to k.

In this case, the condition (5.4.4) of lemma 5.4.2 holds, that is, $d_i^2 = 1$ for all i = 1,2,...,m.

If G is unbalanced, then there exists a cycle σ whose sign is -1. Therefore, we have

$$(-1) \cdot c(d_1,d_2,\ldots,d_m : G) \quad c(d_1,d_2,\ldots,d_m : G). \qquad (5.4.12)$$

This means that

$$c(d_1,d_2,\ldots,d_m : G) \quad 0. \qquad (5.4.13)$$

If $c(d_1,d_2,\ldots,d_m : G) = k$, then every cycle of G has a positive sign. Therefore, G is balanced.

If $c(d_1,d_2,\ldots,d_m : G) \neq k$, then there exists at least one cycle whose sign is equal to -1. Therefore, G is unbalanced. □

Using the above theorem, the calculation of the probability of balance of a graph G under the assumption that the signs of lines are stochastically independent is as follows (cf. Katai, Iwai et al. [1976d] and Katai & Iwai[1978b]).

THEOREM 5.4.2 : Let p_i be the probability that the sign d_i of the line x_i is equal to +1 for i 1,2,...,m. If the signs of lines in G are stochastically independent, then $P_b(p_1,p_2,\ldots,p_m : G)$, the probability of balance of the graph G, is as follows:

$$P_b(p_1,p_2,\ldots,p_m : G) = c(2p_1-1,2p_2-1,\ldots,2p_m-1 : G).k^{-1}. \quad (5.4.14)$$

PROOF : The next equality is a direct consequence of theorem 5.4.1.

$$\begin{aligned} P_b(p_1,p_2,\ldots,p_m : G) &= E[c(d_1,d_2,\ldots,d_m : G)\cdot k^{-1}] \\ &= E[c(d_1,d_2,\ldots,d_m : G]\cdot k^{-1} \\ &= (1 + \sum_{i=1}^{k-1} E[d_{i_1}\cdot d_{i_2}\cdot\cdots\cdot d_{i_{m_i}}])\cdot k^{-1}. \end{aligned} \quad (5.4.15)$$

From the assumption that the random variables $d_1,d_2,\ldots$ and d_m are stochastically independent, we have

$$E[d_{i_1}\cdot d_{i_2}\cdot\cdots\cdot d_{i_{m_i}}] = E[d_{i_1}]\cdot E[d_{i_2}]\cdot\cdots\cdot E[d_{i_{m_i}}]. \quad (5.4.16)$$

According to the probability law of d_i.

$$E[d_i] = 1\cdot p_i + (-1)\cdot(1-p_i) \quad 2p_i \quad 1 \quad \text{for } i = 1,2,\cdots,m. \quad (5.4.17)$$

Substitution of these two formulas (5.4.16) and (5.4.17) to (5.4.15) verifies the theorem. □

As stated in the proof of theorem 5.4.1, the probability of the balance of a graph G is equal to the probability that all of the cycles in G have positive signs. The signs of cycles in G are not stochastically independent. The relation between the probability of balance of G and the probability law of the sign of each cycle in G is given by the following corollary.

COROLLARY 5.4.1 :

$$P_b(p_1,p_2,\cdots,p_m : G) = \frac{2}{k} \cdot \sum_{i=0}^{k-1} \text{Prob.}(s(\sigma_i) = 1) \quad 1, \tag{5.4.18}$$

where $s(\sigma_i)$ is the sign of cycle σ_i for i 0,1,$\cdots$,k-1.

PROOF : As stated in the proof of theorem 5.4.2,

$$P_b(p_1,p_2,\cdots,p_m ; G) = \Big(\sum_{i=0}^{k-1} E[s(\sigma_i)] \Big) \cdot k^{-1} . \tag{5.4.19}$$

The value $s(\sigma_i)$ is +1 or -1, therefore

$$E[s(\sigma_i)] = 1 \cdot \text{Prob.}(s(\sigma_i) = 1) + (-1) \cdot \text{Prob.}(s(\sigma_i) \quad -1)$$

$$2 \cdot \text{Prob.}(s(\sigma_i) = 1) \quad 1 \qquad \text{for } i = 0,1,\cdots,k. \tag{5.4.20}$$

Substituting the above formulas to (5.4.19) verifies the corollary. □

The following corollary is a direct consequence of theorem 5.4.2.

COROLLARY 5.4.2 : For the case where $p_1 \quad p_2 = \cdots = p_m \quad p$, the probability p_b is calculated by the formula that

$$P_b(p,p,\cdots,p : G) = \frac{1}{k} \cdot \sum_{j=0}^{m} c_j (2p - 1)^j , \tag{5.4.21}$$

where c_j is the number of cycles whose length is equal to j. Particularly, when p is equal to $\frac{1}{2}$,

$$P_b(\tfrac{1}{2}, \tfrac{1}{2},\cdots,\tfrac{1}{2} : G) = \frac{1}{k} . \tag{5.4.22}$$

Fig. 5.4.2 shows some examples.

In the remaining part of this section, the interdependence among the balance of graphs (groups) is discussed (cf. Katai, Iwai, et al.[1976d] and Katai & Iwai[1978b]).

As aforementioned, a graph G can be decomposed into its block components such that each circuit of G is contained in exactly one block component of G. If the number of block components of a graph G is equal to t, then k, the number of cycles contained in G, is equal to 2^{m-n+t} where m and n are the numbers of the lines and the vertices of G, respectively. Therefore, equation (5.4.22) in corollary 5.4.2 is written as

COROLLARY 5.4.2' :

$$P_b(\frac{1}{2}, \frac{1}{2}, \cdots, \frac{1}{2} : G) = \frac{1}{k} \quad 2^{n-m-t}. \tag{5.4.22'}$$

If a graph G is decomposed into its block components, then the cycle polynomial of G is also factorized as follows:

LEMMA 5.4.3 : Let $B_1, B_2, \cdots$ and B_t be the block components of G. Then,

$$c(x_1, x_2, \cdots, x_m : G) = c(x_1, x_2, \cdots, x_m : B_1) \cdot c(x_1, x_2, \cdots, x_m : B_2) \cdot \cdots$$
$$\cdots \cdot c(x_1, x_2, \cdots, x_m : B_t). \tag{5.4.23}$$

PROOF : Let $\overline{\Sigma} \quad \{\overline{\sigma}_1, \overline{\sigma}_2, \cdots, \overline{\sigma}_u\}$ be a set of fundamental cycles of G. As aforementioned, we can assume that the cycle $\overline{\sigma}_i$ is a circuit of G for i = 1,2, $\cdots$,u. From the definition of block components, each circuit $\overline{\sigma}_i$ is contained in exactly one block component of G.

Let $\overline{\Sigma}_i$ be the subset of $\overline{\Sigma}$ whose elements are contained in block component B_i for i = 1,2,$\cdots$,t. Then we have

$$\overline{\Sigma} \quad \bigcup_{i=1}^{t} \overline{\Sigma}_i \ , \ \overline{\Sigma}_i \cap \overline{\Sigma}_j = \phi \qquad \text{for } i \neq j \ . \tag{5.4.24}$$

It is easy to show that $\overline{\Sigma}_i$ constitute a set of fundamental cycles of B_i for i 1, 2, $\cdots$, t, because if there is a cycle σ of B_i which cannot be represented by a product of circuits contained in $\overline{\Sigma}_i$, then it cannot be represented by a product of circuits contained in $\overline{\Sigma}$, which contradicts the assumption that $\overline{\Sigma}$ is a set of fundamental cycles of G. Therefore, from equality (5.4.5),

$$c(x_1, x_2, \cdots, x_m : B_i) = \prod_{\overline{\sigma} \in \overline{\Sigma}} (1 + \overline{\sigma}) \qquad \text{for } i = 1,2,\cdots,t. \tag{5.4.25}$$

Taking note of equalities (5.4.5), (5.4.24) and (5.4.25), we have

$$c(x_1, x_2, \cdots, x_m : G) = \prod_{i=1}^{t} c(x_1, x_2, \cdots, x_m : B_i). \tag{5.4.26}$$

The above equality yields the following statement which is equivalent to lemma 5.3.2 in Section 5.3.3.1.

$$c(d_1, d_2, \cdots, d_m : G) \neq 0 \text{ iff } c(d_1, d_2, \cdots, d_m : B_i) \neq 0$$
$$\text{for all } i = 1,2,\cdots,t. \tag{5.4.27}$$

The next theorem is a direct consequence of the above lemma.

THEOREM 5.4.3 : Let $B_1, B_2, \cdots B_t$ be the block components of G, then

$$P_b(p_1,p_2,\cdots,p_m : G) \quad P_b(p_1,p_2,\cdots,p_m : B_1)\cdot P_b(p_1,p_2,\cdots,p_m : B_2)\cdot \cdots \cdot P_b(p_1,p_2,\cdots p_m : B_t). \tag{5.4.28}$$

The effect on the probability of balance of a graph G caused by the removal of a line from G is represented by the next theorem.

THEOREM 5.4.4 : Let x_i be a line of G. Then, if x_i is not contained in any cycle of G, the probability of balance is not affected by the removal of x_i, i.e.,

$$P_b(p_1,p_2,\cdot\cdot,p_{i-1},p_{i+1},\cdots,p_m : G - x_i) = P_b(p_1,p_2,\cdots,p_{i-1},p_i,p_{i+1},\cdots,p_m : G). \tag{5.4.29}$$

If x_i is contained in a cycle of G, then

$$P_b(p_1,p_2,\cdots,p_{i-1},p_{i+1},\cdot\cdot,p_m : G - x_i) = 2\cdot P_b(p_1,p_2,\cdots,p_{i-1},\frac{1}{2},p_{i+1},\cdots,p_m : G) \tag{5.4.30}$$

PROOF : Let Σ_i be a subset of $\Sigma = \{\sigma;\ \sigma \text{ is a cycle of G}\}$ as follows:

$$\Sigma_i = \{\sigma;\ \sigma \text{ contains the line } x_i\}. \tag{5.4.31}$$

It is clear that

$$\text{if } \Sigma_i = \phi, \text{ then } \Sigma - \Sigma_i \quad \Sigma. \tag{5.4.32}$$

Assume that $\Sigma_i \neq \phi$. Let σ' be a cycle contained in Σ_i. Then, from equality (5.4.8), we have

$$\{\sigma'\cdot\sigma\ ;\ \sigma \in \Sigma\} = \Sigma \tag{5.4.33}$$

It is clear that

$$\text{if } \sigma \in \Sigma_i \text{ then } \sigma'\cdot\sigma \in \Sigma \quad \Sigma_i \tag{5.4.34}$$

and

$$\text{if } \sigma \in \Sigma \quad \Sigma_i \text{ then } \sigma'\cdot\sigma \in \Sigma_i. \tag{5.4.35}$$

That is,

$$\{\sigma'\cdot\sigma\ ;\ \sigma \in \Sigma_i\} = \Sigma \quad \Sigma_i. \tag{5.4.36}$$

Taking note of (5.4.33) and (5.4.36), we have

$$\#(\Sigma - \Sigma_i) = \#(\Sigma_i). \tag{5.4.37}$$

Therefore, the next equalities are valid.

$$\#(\Sigma - \Sigma_i) = \#(\Sigma) = k \qquad \text{if } \Sigma_i \quad \phi, \tag{5.4.38}$$

$$\#(\Sigma - \Sigma_i) \quad \frac{1}{2}k \qquad \text{if } \Sigma_i \neq \phi. \tag{5.4.39}$$

If $p_i \quad \frac{1}{2}$, then the probability of balance of a graph G is as follows:

$$P_b(p_1,p_2,\cdots,p_{i-1},\frac{1}{2},p_{i+1},\cdots,p_m : G)$$

$$c(2p_1 \quad 1,\ 2p_2 \quad 1,\cdots,\ 2p_{i-1} \quad 1,\ 0,\ 2p_{i+1} \quad 1,\cdots,\ 2p_m \quad 1 : G)\cdot k^{-1}. \tag{5.4.40}$$

On the other hand, we have

$$P_b(p_1,p_2,\cdots,p_{i-1},p_{i+1},\cdots,p_m : G - x_i)$$

$$\begin{cases} c(2p_1-1,2p_2-1,\cdots,2p_{i-1}-1,2p_{i+1}-1,\cdots,2p_m-1 : G-x_i)\cdot k^{-1} \\ \qquad\qquad \text{if } \Sigma_i = \phi, \qquad (5.4.41) \\ c(2p_1-1,2p_2-1,\cdots,2p_{i-1}-1,2p_{i+1}-1,\cdots,2p_m-1 : G-x_i)\cdot(\frac{1}{2}k)^{-1} \\ \qquad\qquad \text{if } \Sigma_i \neq \phi. \qquad (5.4.42) \end{cases}$$

From the definition of cycle polynomial, it is easy to see that

$$c(x_1,x_2,\cdots,x_{i-1},0,x_{i+1},\cdots,x_m : G)$$

$$= c(x_1,x_2,\cdots,x_{i-1},x_{i+1},\cdots,x_m : G-x_i). \tag{5.4.43}$$

Taking note of (5.4.40), (5.4.41), (5.4.42), and (5.4.43), the theorem is verified. □

Let G_1 and G_2 be two arbitrary subgraphs of G. Then the following discussions are valid.

THEOREM 5.4.5 : If the conditon that

$$G_2 \supset G_1 \quad \text{or} \quad G_2 \subset G_1 \tag{5.4.44}$$

hold, then the next inequality is valid for arbitrary probability assingnments for G.

$$\text{Prob.}(G_2 \text{ is balanced} \mid G_1 \text{ is balanced}) \geq \text{Prob.}(G_2 \text{ is balanced}). \tag{5.4.45}$$

PROOF : According to the definition of conditional probability and to the fact that any subgraph of a balanced graph is also balanced, the following equalities hold.

$$\begin{aligned} &\text{Prob.}(G_2 \text{ is balanced} \mid G_1 \text{ is balanced}) \\ &\triangleq \frac{\text{Prob.}(G_1 \text{ and } G_2 \text{ are balanced})}{\text{Prob.}(G_1 \text{ is balanced})} \\ &= \begin{cases} \dfrac{\text{Prob.}(G_2 \text{ is balanced})}{\text{Prob.}(G_1 \text{ is balanced})} & \text{if } G_2 \supset G_1, \qquad (5.4.46) \\ 1 & \text{if } G_2 \subset G_1. \qquad (5.4.47) \end{cases} \end{aligned}$$

Therefore, inequality (5.4.45) is valid. □

However, if condition (5.4.44) does not hold, then (5.4.45) is not necessarily valid. For example, in Fig. 5.4.3, let G_1 and G_2 be as follows:

$$G_1 = (\{v_1,v_2,v_3,v_4\},\{x_1,x_2,x_3,x_4\}),\ G_2 \quad (\{v_2,v_3,v_4\},\{x_3,x_4,x_5\}). \tag{5.4.48}$$

And let p_i be equal to p for $i = 1,2,\cdots,5$. Then, the probabilities are

$$\text{Prob.}(G_2 \text{ is balanced} \mid G_1 \text{ is balanced}) \quad \frac{(1 + (2p \quad 1)^3)^2}{1 + (2p - 1)^4}, \tag{5.4.49}$$

$$\text{Prob.}(G_2 \text{ is balanced}) \quad 1 + (2p - 1)^3. \tag{5.4.50}$$

Therefore ,

$$\text{if } p \leq \frac{1}{2}, \text{ then } P_b(G_2 \text{ is balanced} \mid G_1 \text{ is balanced}) \leq P_b(G_2 \text{ is balanced}), \tag{5.4.51}$$

$$\text{if } p \geq \frac{1}{2}, \text{ then } P_b(G_2 \text{ is balanced} \mid G_1 \text{ is balanced}) \geq P_b(G_2 \text{ is balanced}). \tag{5.4.52}$$

The above inequality (5.4.52) generally holds by the following statement.

THEOREM 5.4.6 : If the assignment of probabilities satisfies the condition that

$$p_i \geq \frac{1}{2} \qquad \text{for all } i = 1,2,\cdots,m \tag{5.4.53}$$

then,

$$P_b(G_2 \text{ is balanced} \mid G_1 \text{ is balanced}) \geq P_b(G_2 \text{ is balanced}) \tag{5.4.54}$$

for arbitrary subgraphs G_1 and G_2.

PROOF : From the definition of conditional probability, we have

$$\text{Prob.}(G_2 \text{ is balanced} \mid G_1 \text{ is balanced}) = \frac{\text{Prob.}(G_1 \text{ and } G_2 \text{ are balanced})}{\text{Prob.}(G_1 \text{ is balanced})}. \tag{5.4.55}$$

Let $\{\sigma_0^1, \sigma_1^1, \sigma_2^1, \cdots, \sigma_{k_1-1}^1\}$ and $\{\sigma_0^2, \sigma_1^2, \cdots, \sigma_{k_2-1}^2\}$ be the sets of the cycles in G_1 and G_2, respectively. Then, from theorem 5.4.1, we have

$$\text{Prob.}(G_1 \text{ and } G_2 \text{ are balanced}) = E[(\sum_{i=0}^{k-1} \sigma_i^1)\cdot k_1^{-1} \cdot (\sum_{j=0}^{k-1} \sigma_j^2)\cdot k_2^{-1}]. \tag{5.4.56}$$

Also, we have

$$\text{Prob.}(G_1 \text{ is balanced})\cdot\text{Prob.}(G_2 \text{ is balanced}) \quad E[(\sum_{i=0}^{k-1} \sigma_i^1)\cdot k_1^{-1}]\cdot E[(\sum_{j=0}^{k-1} \sigma_j^2)\cdot k_2^{-1}]. \tag{5.4.57}$$

Let $\sigma = \prod_{i\in I} x_i$ and $\sigma' = \prod_{i\in I'} x_i$ be arbitrary two cycles of G. Then, we have

$$E[\sigma\cdot\sigma'] - E[\sigma]\cdot E[\sigma'] \quad \prod_{i\in I\cup I' \; I\cap I'} (2p_i - 1). \prod_{i\in I\cap I'} (1 \quad (2p_i - 1)^2). \tag{5.4.58}$$

Therefore, if condition (5.4.53) holds, then the right-hand side of the above equality is nonnegative, i.e.,

$$E[\sigma\cdot\sigma'] \geq E[\sigma]\cdot E[\sigma'] \tag{5.4.59}$$

This leads to the following inequality.

$$E[(\sum_{i=0}^{k-1} \sigma_i^1)\cdot k_1 \; {}^1\cdot(\sum_{j=0}^{k-1} \sigma_j^2)\cdot k_2^{-1}] \geq E[(\sum_{i=0}^{k-1} \sigma_i^1)\cdot k_1^{-1}]\cdot E[(\sum_{j=0}^{k-1} \sigma_j^2)\cdot k_2^{-1}]. \tag{5.4.60}$$

Therefore, inequality (5.4.54) is valid. □

In the next section, we discuss the general case where the assumption of independence does not hold.

5.4.2 Classification of Types of Balance by Finite-state Systems Theoretical Aspect

As aforementioned in the introduction, interpersonal relationships are sometimes determined by the members' opinions or attitudes about certain social arguments, beliefes or norms. Suppose that the opinions or attitudes of each member are categorized to finite items (internal states). If the signs of the interpersonal relations are consistent with this categorization, then the following interpersonal group structure has significant meaning in examining the actual situations of social groups (cf. Katai, Iwai, et al.[1975a]and Katai & Iwai[1978c]).

DEFINITION 5.4.2 : *A group with a relational structure (a group w.r.s.)* is an ordered 4-tuple

$$R = (V,X,\{S(v_i)\}_{v_i \in V},\{R_{v_i v_j}\}_{\{v_i,v_j\} \in X}), \tag{5.4.61}$$

where (1) $V = \{v_1,v_2,\cdots,v_n\}$ is a finite nonempty set, referred to as the set of members (in the group); n represents the number of members. (2) X is contained in the product set $V \otimes V$ of distinct unordered pairs of the elements of V, called the set of direct relations (in the group). (3) We call the graph $G = (V,X)$ the graph structure of R and denote it by $G(R)$. (4) $S(v_i)$ is a finite nonempty set representing the set of internal states of the member v_i. (5) $R_{v_i v_j}$ is a function mapping the product set $S(v_i) \times S(v_j)$ into the set $\{+1,-1\}$, where $\{v_i,v_j\} \in X$ and also $R_{v_i v_j}$ is symmetric in the following sense:

$$R_{v_i v_j}(s_i,s_j) = R_{v_j v_i}(s_j,s_i) \qquad \text{for all } s_i \in S(v_i), \text{ all } s_j \in S(v_j). \tag{5.4.62}$$

The value $R_{v_i v_j}(s_i,s_j)$ represents the interpersonal relation (+1 or -1) between the members v_i and v_j when v_i and v_j teke their internal states as s_i and s_j, respectively. We call this function $R_{v_i v_j}$ the *relation function* of members v_i and v_j.

Fig. 5.4.4 shows an example with members v_1,v_2 and v_3. If they take their internal states as s_1^1,s_2^1 and s_3^1, respectively, then their interpersonal rela-

tions are +1, +1 and +1, respectively and the group is balanced. If the member v_2 changes his internal state (his opinion) from s_2^1 to s_2^2, then the relations become -1, +1 and +1, and the group becomes unbalanced.

In the context of this group structure, the properties of balance are classified as follows:

DEFINITION 5.4.3 : A group w.r.s $\mathcal{R}$ is said to be *totally balanced* iff the group is balanced no matter how each member in the group takes any of his internal states. A group w.r.s.$\mathcal{R}$ is said to be *totally unbalanced* iff the group is unbalanced no matter how each member in the group takes any of his internal states. A group w.r.s.$\mathcal{R}$ is said to be *partially balanced* iff the above two requirements do not hold, that is, the balance of the group depends on the internal states taken by the members.

In Fig.5.4.5, these balance types are shown in the case of a group with three members similar to Fig.5.4.4.

A subgroup of a group w.r.s is defined as follows.

DEFINITION 5.4.4 : Let $\mathcal{R} = (V,X,\{S(v_i)\}_{v_i\in V},\{R_{v_iv_j}\}_{\{v_i,v_j\}\in X})$ be a group w.r.s. and $\overline{G}$ $(\overline{V},\overline{X})$ be a subgraph of $G(\mathcal{R}) = (V,X)$. The *restriction* of $\mathcal{R}$ to $\overline{G}$ is a group w.r.s. denoted by $\mathcal{R}|_{\overline{G}}$ is as follows:

$$\mathcal{R}|_{\overline{G}} \triangleq (\overline{V},\overline{X},\{S(v_i)\}_{v_i\in\overline{V}},\{R_{v_iv_j}\}_{\{v_i,v_j\}\in\overline{X}}). \tag{5.4.63}$$

The discussion about the balance types of a group $\mathcal{R}$ is simplified using block decomposition of the graph structure $G(\mathcal{R})$ of $\mathcal{R}$ (cf. Katai & Iwai[1978b]).

THEOREM 5.4.7 : For any group w.r.s. $\mathcal{R}$, the following statements hold, where $B_1,B_2,\cdots$ and B_t are the block components of $G(\mathcal{R})$. $\mathcal{R}$ is tatally balanced if all the block components of $G(\mathcal{R})$ are totally balanced, i.e., $\mathcal{R}|_{B_1}$, $\mathcal{R}|_{B_2},\cdots$ and $\mathcal{R}|_{B_t}$ are totally balanced.

PROOF : As stated in lemma 5.3.2, an s-graph is balanced iff all of its block components are balanced. Therefore, $\mathcal{R}$ is totally balanced iff $\mathcal{R}|_{B_1}$, $\mathcal{R}|_{B_2},\cdots$ and $\mathcal{R}|_{B_t}$ are totally balanced. □

For totally unbalanced or partially balanced cases, only the sufficient or necessary conditions can be shown as follows:

THEOREM 5.4.8 :

(1) $\mathcal{R}$ is totally unbalanced if at least one of the block components $\mathcal{R}|_{B_1}$, $\mathcal{R}|_{B_2}, \cdots$ and $\mathcal{R}|_{B_t}$ is tatally unbalanced.

(2) $\mathcal{R}$ is partially balanced only if none of the block components $\mathcal{R}|_{B_1}$, $\mathcal{R}|_{B_2}, \cdots$ and $\mathcal{R}|_{B_t}$ is totally unbalanced and at least one of them is partially balanced.

PROOF : The statement (1) is clear because any group w.r.s. $\mathcal{R}$ is totally unbalanced if its subgroup is totally unbalanced.

If $\mathcal{R}$ is partially balanced, then there exist internal states $s_1, s_2, \cdots, s_m$ such that $s_i \in S(v_i)$ for $i = 1,2,\cdots,m$ and the resultant s-graph is balanced. Therefore, if $\mathcal{R}$ is partially balanced then none of $\mathcal{R}|_{B_1}, \mathcal{R}|_{B_2}, \cdots$ and $\mathcal{R}|_{B_t}$ are not totally unbalanced and not all of them are totally balanced. □

According to theorem 5.4.7, for the characterization of totally balanced groups, it is sufficient to consider the case where $G(\mathcal{R})$ is nonseparable.

A graph consisting of only two vertices (one line) is considered (vacousely) balanced. We call a nonseparable graph consisting of at least three vertices as *strictly nonseparable graph.*

If two members v_i and v_j have a direct relation, that is, $\{v_i, v_j\} \in X$, then the function $R_{v_i v_j}$ is a mapping from $S(v_i) \times S(v_j)$ into the set $\{+1, -1\}$. Therefore, the inverse function (mapping) $R^{-1}_{v_i v_j}$ produces a partition $\{R^{-1}_{v_i v_j}(+1), R^{-1}_{v_i v_j}(-1)\}$ on the set $S(v_i) \times S(v_j)$, i.e.,

$$\left.\begin{aligned} R^{-1}_{v_i v_j}(+1) \cup R^{-1}_{v_i v_j}(-1) &= S(v_i) \times S(v_j) \\ R^{-1}_{v_i v_j}(+1) \cap R^{-1}_{v_i v_j}(-1) &= \Phi \end{aligned}\right\} . \qquad (5.4.64)$$

We denote by $\mathbb{P}_i^j = \{S(v_i)_j^+ , S(v_i)_j^-\}$ any partition of $S(v_i)$, where suffix j corresponds to the member v_j that has a direct relation with v_i. Let $\mathbf{P} = \{S\}$ and $\mathbf{P}' = \{S'\}$ be two partitions on the same set We call $\mathbf{P}$ a refinement of $\mathbf{P}'$ or $\mathbf{P}$ is finer than $\mathbf{P}'$, if every set in $\mathbf{P}$ is contained in a set S' of $\mathbf{P}'$. In other words, each set S' in $\mathbf{P}'$ is a union of some of the sets in $\mathbf{P}$, and $\mathbf{P}'$ is said to be coarser than $\mathbf{P}$.

LEMMA 5.4.4 : Let $\mathcal{R}$ be a group w.r.s. such that $G(\mathcal{R})$ is strictly nonseparable. If $\mathcal{R}$ is totally balanced, then for any pair of members v_i and v_j such that $\{v_i, v_j\} \in X$, there exist partitions $\mathbb{P}_i^j$ and $\mathbb{P}_j^i$ on $S(v_i)$ and $S(v_j)$, respectively, such that the partition $\{R^{-1}_{v_i v_j}(+1), R^{-1}_{v_i v_j}(-1)\}$ is coarser than

the partition $\mathbb{P}_i^j \times \mathbb{P}_j^i$, i.e. the product partition $\{S(v_i)_j^+ \times S(v_j)_i^+,\ S(v_i)_j^+ \times S(v_j)_i^-,\ S(v_i)_j^- \times S(v_j)_i^+,\ S(v_i)_j^- \times S(v_j)_i^-\}$ on the set $S(v_i) \times S(v_j)$.

PROOF : From the definition of strictly nonseparable graphs, the line $x = \{v_i, v_j\}$ is contained in a circuit σ_{ij} of $G(R)$. Let $v_{k_1}, v_{k_2}, \cdots$ and v_{k_r} be the vertices contained in σ_{ij}. If the group R is totally balanced, then for any internal state $s_{k_q} \in S(v_{k_q})$ for $q = 1,2,\cdots,r$, the signed circuit $\sigma_{ij}(s_{k_1}, s_{k_2}, \cdots, s_{k_r})$ has positive sign where $\sigma_{ij}(s_{k_1}, s_{k_2}, \cdots, s_{k_r})$ is constructed from σ_{ij} by attaching the sign $R_{v_{k_q} v_{k_{q+1}}}(s_{k_q}, s_{k_{q+1}})$ to the line $\{v_{k_q}, v_{k_{q+1}}\}$ for $q = 1,2, \cdots, r$ (for convenience $v_{k_{r+1}}$ is set to be v_{k_1}). Let v_a and v_b be the vertices adjacent to v_i and v_j in the circuit σ_{ij}, that is, the lines $\{v_a, v_i\}$ and $\{v_j, v_b\}$ are contained in σ_{ij} (see Fig.5.4.6). Let $s_{k_1}, s_{k_2}, \cdots, s_a, s_b, \cdots$ and s_{k_r} be fixed. For any internal states $s_i \in S(v_i)$ and $s_j \in S(v_j)$, the following equalty holds.

$$\text{the sign of } \sigma_{ij}(s_{k_1}, s_{k_2}, \cdots, s_a, s_i, s_j, s_b, \cdots, s_{k_r}) = +1. \tag{5.4.65}$$

That is, for fixed $s_a \in S(v_a)$ and $s_b \in S(v_b)$, the following function $f(s_i, s_j)$ has a constant sign $h(s_a, s_b)$ (+1 or -1), i.e.,

$$f(s_i, s_j) \triangleq R_{v_a v_i}(s_a, s_i) \cdot R_{v_i v_j}(s_i, s_j) \cdot R_{v_j v_b}(s_j, s_b) \equiv h(s_a, s_b), \tag{5.4.66}$$

In the other words,

$$R_{v_i v_j}(s_i, s_j) = h(s_a, s_b) \cdot R_{v_a v_i}(s_a, s_i) \cdot R_{v_j v_b}(s_j, s_b). \tag{5.4.67}$$

Let $S(v_i)_j^+$ and $S(v_i)_j$ be the subsets of $S(v_i)$ as follows:

$$\left.\begin{aligned} S(v_i)_j^+ &= R_{v_a v_i}(s_a, \cdot)^{-1}(+1) \\ S(v_i)_j &= R_{v_a v_i}(s_a, \cdot)^{-1}(-1) \end{aligned}\right\}. \tag{5.4.68}$$

Similary, we set a partition on $S(v_j)$ as follows:

$$\left.\begin{aligned} S(v_j)_i^{+1} &= R_{v_j v_b}(\cdot, s_b)^{-1}(+1) \\ S(v_j)_i^{-1} &\quad R_{v_j v_b}(\cdot, s_b)^{-1}(-1) \end{aligned}\right\}. \tag{5.4.69}$$

Then, from equation (5.4.67), the partitions $\mathbb{P}_i^j = \{S(v_i)_j^+, S(v_i)_j^-\}$ and

$\mathbf{P}_j^i = \{S(v_j)_i^+, S(v_j)_i^-\}$ satisfy the condition of the theorem. □

The next corollary is a direct consequence of the above lemma.

COROLLARY 5.4.3 : Let $\mathbb{P}_i = \{S(v_i)^+, S(v_i)^-\}$ and $\mathbf{P}_j = \{S(v_j)^+, S(v_j)^-\}$ denote arbitrary partitions on $S(v_i)$ and $S(v_j)$, respectively. If v_i and v_j have a direct relation $R_{v_iv_j}$ such that the partition $\{R_{v_iv_j}^{-1}(+1), R_{v_iv_j}^{-1}(-1)\}$ is not coarser than any product partition $\mathbb{P}_i \times \mathbb{P}_j$, then any group $\mathcal{R}$ containing v_i and v_j is not totally balanced when its graph structure $G(\mathcal{R})$ is strictly nonseparable.

For example, the members v_1 and v_2 in (1) and (2) in Fig. 5.4.7 satisfy the condition of the above corollary. The group $\mathcal{R}_1$ is not totally balanced because the graph structure $G(\mathcal{R}_1)$ is strictly nonseparable However, it is easy to see that the group $\mathcal{R}_2$ in (2) is totally balanced. Therefore, the condition of strict nonseparability is essential.

In the above lemma 5.4.4, the partition $\mathbb{P}_i^j$ on $S(v_i)$ does not depend on the member v_j, i.e., the following theorem is verified.

THEOREM 5.4.9 : Let $\mathcal{R}$ be the same as stated in lemma 5.4.4. Then $\mathcal{R}$ is totally balanced iff there exists a partition $\{S(v_i)^+, S(v_i)^-\}$ on $S(v_i)$ for i 1, 2, ···, m such that the following conditions are satisfied.

$$R_{v_iv_j}(s_i,s_j) = +1 \quad \text{if } s_i \in S(v_i)^+,\ s_j \in S(v_j)^+ \text{ or } s_i \in S(v_i),\ s_j \in S(v_j) \tag{5.4.70}$$

$$R_{v_iv_j}(s_i,s_j) = -1 \quad \text{if } s_i \in S(v_i)^+,\ s_j \in S(v_j)^- \text{ or } s_i \in S(v_i)^-,\ s_j \in S(v_j)^+ \tag{5.4.71}$$

PROOF : Assume that $\mathcal{R}$ is totally balanced. Let s_c and s_d be the internal states of v_c and v_d, respectively, as shown in Fig.5.4.8. According to the same method used for the derivation of equation (5.4.66), all the paths joining s_c and s_d have the same sign $h(s_c, s_d)$. Let s_c be fixed and assigned to +1. Then we can assign +1 or -1 to any internal state s_d according to the sign $h(s_d) = h(s_c, s_d)$. From the definition of total balance that every circuit has the sign +1, the following equation holds for any v_i and v_j such that $\{v_i, v_j\} \in X$ (see Fig. 5.4.9).

$$h(s_i)\cdot R_{v_iv_j}(s_i,s_j)\cdot h(s_j) \quad +1 \quad \text{for all } s_i \in S(v_i), \text{ all } s_j \in S(v_j). \tag{5.4.72}$$

That is,

$$R_{v_i v_j}(s_i, s_j) = h(s_i) \cdot h(s_j). \tag{5.4.73}$$

Let the sets $S(v_i)^+$ and $S(v_j)$ be as follows for i = 1,2,··,m.

$$\left.\begin{aligned} S(v_i)^+ &= \{s_i \in S(v_i) \,;\, h(s_i) = +1\} \\ S(v_i)^- &= \{s_i \in S(v_i) \,;\, h(s_i) = -1\} \end{aligned}\right\} . \tag{5.4.74}$$

By equation (5.4.73), the partitions $S(v_i)^+$, $S(v_i)^-$ (i = 1,2,···,n) satisfy the condition of the theorem.

To prove the sufficience, we use the notation as follows.

$$\left.\begin{aligned} h(s_i) &= +1 \quad \text{if } s_i \in S(v_i)^+ \\ h(s_i) &= -1 \quad \text{if } s_i \in S(v_i)^- \end{aligned}\right\} . \tag{5.4.75}$$

Then, by conditions (5.4.70) and (5.4.71),

$$R_{v_i v_j}(s_i, s_j) = h(s_i) \cdot h(s_j) \quad \text{for all } s_i \in S(v_i), \text{ all } s_j \in S(v_j) \tag{5.4.76}$$

Let $\sigma(s_{k_1}, s_{k_2}, \cdots, s_{k_r})$ be a circuit in R. The sign of σ is defined by

$$\begin{aligned} \text{the sign of } \sigma \quad & R_{v_{k_1} v_{k_2}}(s_{k_1}, s_{k_2}) \cdot R_{v_{k_2} v_{k_3}}(s_{k_2}, s_{k_3}) \cdot \cdots \cdot \\ & R_{v_{k_r} v_{k_1}}(s_{k_r}, s_{k_1}) \\ = \; & h(s_{k_1}) \cdot h(s_{k_2})^2 \cdot \cdots \cdot h(s_{k_r})^2 \cdot h(s_{k_1}) = +1 \end{aligned} \tag{5.4.77}$$

Therefore, R is totally balanced. □

THEOREM 5.4.10 : The partition $\{S(v_i)^+, S(v_i)^-\}$ in theorem 5.4.9 is unique for i = 1,2,···,n.

PROOF : Let v_i and v_j be the vertices adjacent to each other, and $\{S_\alpha(v_i)^+, S_\alpha(v_i)^-\}$ and $\{S_\beta(v_i)^+, S_\beta(v_i)^-\}$ be partitions on $S(v_i)$ for i = 1,2,···,n satisfying the condition of theorem 5.4.9. Then, from equations (5.4.70), (5.4.71) and (5.4.76),

$$h_\alpha(s_i) \cdot h_\alpha(s_j) = h_\beta(s_i) \cdot h_\beta(s_j) \quad (\; R_{v_i v_j}(s_i, s_j) \;), \tag{5.4.78}$$

where h_α and h_β are the signs defined as similar to (5.4.75). Multiplying $h_\alpha(s_j)\cdot h_\beta(s_i)$ to the both sides of (5.4.78), we have

$$h_\alpha(s_i)\cdot h_\beta(s_i) \quad h_\alpha(s_j)\cdot h_\beta(s_j), \quad \text{for all } s_i \in S(v_i), \text{ all } s_j \in S(v_j). \tag{5.4.79}$$

The left side is a function of s_i, and the other side is a function of s_j. Therefore, we have

$$h_\alpha(s_i) \equiv h_\beta(s_i) \text{ or } h_\alpha(s_i) \equiv -h_\beta(s_i). \tag{5.4.80}$$

That is,

$$\{S_\alpha(v_i)^+, S_\alpha(v_i)^-\} - \{S_\beta(v_i)^+, S_\beta(v_i)^-\}. \tag{5.4.81}$$ □

A graph G = (V,X) is a complete graph when X V ⊗ V. The above theorem means that the members in a totally balanced group have an enough consensus about their opinions in the following sense.

COROLLARY 5.4.4 : Let $\mathcal{R}$ be a totally balanced group such that $G(\mathcal{R})$ is strictly nonseparable. Then, we can extend the group $\mathcal{R}$ to a totally balanced group $\mathcal{R}'$ whose graph structure $G(\mathcal{R}')$is a complete graph, by adding direct relations between the members not adjacent to each other in the original graph $G(\mathcal{R})$.

PROOF : When a line $\{v_i,v_j\}$ is not contained in $G(\mathcal{R})$, we make the function $R_{v_iv_j}$ according to the equations (5.4.70) and (5.4.71) in theorem 5.4.9. The resultant group $\mathcal{R}'$ satisfies the conditions of theorem 5.4.9, so it is totally balanced. □

For example, in Fig.5.4.10, the group $\mathcal{R}_1$ is totally balanced and it can be extended to the totally balanced group $\mathcal{R}_2$ by adding relations $R_{v_1v_3}$ and $R_{v_2v_4}$.

Theorems 5.4.9 and 5.4.10 completely characterize totally balanced groups.

For totally unbalanced or partially balanced groups, however, the characterizations are not so easy. We can show the necessary and sufficient condition of total unbalance only for the case when $G(\mathcal{R})$ is composed of only one circuit as follows (cf. Katai & Iwai[1978b]).

THEOREM 5.4.11 : Let $\mathcal{R}$ be a group w.r.s. such that $G(\mathcal{R})$ contains exactly one circuit σ and $v_{k_1},v_{k_2},\cdots,v_{k_r}$ be the vertices contained in the circuit. Namely, the circuit σ is composed of the lines $\{v_{k_1},v_{k_2}\}$, $\{v_{k_2},v_{k_3}\},\cdots$ and $\{v_{k_r},v_{k_1}\}$. Then $\mathcal{R}$ is totally unbalanced iff there exist partitions $\{S(v_{k_i})^+,$

$S(v_{k_i})^-\}$ on $S(v_{k_i})$ for $i = 1,2,\cdots,r$ and also a member v_{k_q} $(1 \le q \le r)$ such that the following conditions are satisfied.

$$\left.\begin{array}{l} s_{k_i} \in S(v_{k_i})^+,\ s_{k_{i+1}} \in S(v_{k_{i+1}})^+ \\ \text{or} \\ s_{k_i} \in S(v_{k_i})^-,\ s_{k_{i+1}} \in S(v_{k_{i+1}})^- \end{array}\right\} \Longrightarrow R_{v_{k_i} v_{k_{i+1}}}(s_{k_i}, s_{k_{i+1}}) = +1$$

$$\left.\begin{array}{l} s_{k_i} \in S(v_{k_i})^+,\ s_{k_{i+1}} \in S(v_{k_{i+1}}) \\ \text{or} \\ s_{k_i} \in S(v_{k_i}),\ s_{k_{i+1}} \in S(v_{k_{i+1}})^+ \end{array}\right\} \Longrightarrow R_{v_{k_i} v_{k_{i+1}}}(s_{k_i}, s_{k_{i+1}}) = -1$$

for $i \neq q$ (for convenience, we set $v_{k_{r+1}} \quad v_{k_1}$) (5.4.82)

and

$$\left.\begin{array}{l} s_{k_q} \in S(v_{k_q})^+,\ s_{k_{q+1}} \in S(v_{k_{q+1}})^+ \\ \text{or} \\ s_{k_q} \in S(v_{k_q})^-,\ s_{k_{q+1}} \in S(v_{k_{q+1}}) \end{array}\right\} \Longrightarrow R_{v_{k_q} v_{k_{q+1}}}(s_{k_q}, s_{k_{q+1}}) = -1$$

$$\left.\begin{array}{l} s_{k_q} \in S(v_{k_q})^+,\ s_{k_{q+1}} \in S(v_{k_{q+1}})^- \\ \text{or} \\ s_{k_q} \in S(v_{k_q})^-,\ s_{k_{q+1}} \in S(v_{k_{q+1}})^+ \end{array}\right\} \Longrightarrow R_{v_{k_q} v_{k_{q+1}}}(s_{k_q}, s_{k_{q+1}}) = +1$$

(5.4.83)

PROOF : Let $B_1, B_2, \cdots$ and B_t be the block components of $G(R)$. Then, from the definition of block components, one of the block components, say B_1, equals σ. The others $B_2, B_3, \cdots$ and B_t are composed of only two vertices. A signed graph without circuits is always balanced, therefore, $R|_{B_2}$, $R|_{B_3}$, $\cdots$ and $R|_{B_t}$ are totally balanced. According to theorem 5.4.8, R is totally unbalanced iff $R|_{B_1}$ is totally unbalanced. Suppose that the conditions (5.4.82) and (5.4.83) are satisfied. Using the notation defined as (5.4.75), the sign of circuit $\sigma(s_{k_1}, s_{k_2}, \cdots, s_{k_r})$ is as follows:

$$\begin{aligned} \text{the sign of } \sigma(s_{k_1}, s_{k_2}, \cdots, s_{k_r}) &= (h(s_{k_1}) \cdot h(s_{k_2})) \cdot (h(s_{k_2}) \cdot h(s_{k_3})) \cdot \cdots \\ &\cdots \cdot (-h(s_{k_q}) \cdot h(s_{k_{q+1}})) \cdot (h(s_{k_{q+1}}) \cdot h(s_{k_{q+2}})) \cdot \cdots \cdot (h(s_{k_r}) \cdot h(s_{k_1})) \\ &-h(s_{k_1})^2 \cdot h(s_{k_2})^2 \cdot \cdots \cdot h(s_{k_r})^2 = -1 \end{aligned} \tag{5.4.84}$$

Therefore, $R|_{B_1}$ is totally unbalanced, i.e., R is totally unbalanced.

If $\mathcal{R}$ is totally unbalanced, i.e., $\mathcal{R}|_{B_1}$ ($= \mathcal{R}|_{\sigma}$) is totally unbalanced, then the following equality holds.

$$\text{The sign of } \sigma(s_{k_1}, s_{k_2}, \cdots, s_{k_r}) = R_{v_{k_1} v_{k_2}}(s_{k_1}, s_{k_2}) \cdot R_{v_{k_2} v_{k_3}}(s_{k_2}, s_{k_3}) \cdot \cdot$$

$$\cdots \cdot R_{v_{k_{r-1}} v_{k_r}}(s_{k_{r-1}}, s_{k_r}) \cdot R_{v_{k_r} v_{k_1}}(s_{k_r}, s_{k_1}) \quad -1$$

$$\text{for all } s_{k_1} \in S(v_{k_1}), \text{ all } s_{k_2} \in S(v_{k_2}), \cdots \text{ and all } s_{k_r} \in S(v_{k_r}).$$

(5.4.85)

Let v_{k_q} be arbitrary one of v_{k_1}, $v_{k_2}, \cdots$ and v_{k_r}. We make a group w.r.s. $\mathcal{R}'$ whose graph structure $G(\mathcal{R}')$ equals to B_1 ($= \sigma$) and the sets of relational functions are defined as follows:

$$R'_{v_{k_i} v_{k_{i+1}}}(s_{k_i}, s_{k_{i+1}}) = R_{v_{k_i} v_{k_{i+1}}}(s_{k_i}, s_{k_{i+1}}),$$

$$\text{for all } i \neq q, \text{ all } s_{k_i} \in S(v_{k_i}), \; s_{k_{i+1}} \in S(v_{k_{i+1}}), \tag{5.4.86}$$

$$R'_{v_{k_q} v_{k_{q+1}}}(s_{k_q}, s_{k_{q+1}}) \quad -R_{v_{k_q} v_{k_{q+1}}}(s_{k_q}, s_{k_{q+1}}),$$

$$\text{for all } s_{k_q} \in S(v_{k_q}), \; s_{k_{q+1}} \in S(v_{k_{q+1}}). \tag{5.4.87}$$

Then, from equation (5.4.87), the sign of the circuit $\sigma'(s_{k_1}, s_{k_2}, \cdots, s_{k_r})$ in the group $\mathcal{R}'$ is given as

$$\text{the sign of } \sigma'(s_{k_1}, s_{k_2}, \cdots, s_{k_r}) \quad +1$$

$$\text{for all } s_{k_1} \in S(v_{k_1}), \; s_{k_2} \in S(v_{k_2}), \cdots \text{ and } s_{k_r} \in S(v_{k_r}).$$

(5.4.88)

Namely, $\mathcal{R}'$ is totally balanced. From theorem 5.4.9, there exists a partition $\{S(v_{k_i})^+, S(v_{k_i})^-\}$ on $S(v_{k_i})$ for $i = 1,2,\cdots,r$ such that the conditions below are satisfied for $i = 1,2,\cdots,r$.

$$R'_{v_{k_i} v_{k_{i+1}}}(s_{k_i}, s_{k_{i+1}}) = +1$$

$$\text{if } s_{k_i} \in S(v_{k_i})^+, s_{k_{i+1}} \in S(v_{k_{i+1}})^+ \text{or } s_{k_i} \in S(v_{k_i})^-, s_{k_{i+1}} \in S(v_{k_{i+1}})^-.$$

(5.4.89)

$$R'_{v_{k_i} v_{k_{i+1}}}(s_{k_i}, s_{k_{i+1}}) = -1$$

$$\text{if } s_{k_i} \in S(v_{k_i})^+, s_{k_{i+1}} \in S(v_{k_{i+1}})^- \text{ or } s_{k_i} \in S(v_{k_i})^-, s_{k_{i+1}} \in S(v_{k_{i+1}})^+. \tag{5.4.90}$$

Therefore, the above partitions satisfy conditions (5.4.82) and (5.4.83) of the theorem. □

For example, in Fig. 5.4.11 (1), the group R is totally unbalanced and the corresponding partitions are indicated by Fig. 5.4.11 (2).

In the above, we have discussed the characterization of the types of social balance in general, i.e., no prescriptions have been posed on the relational functions R_{ij}'s. In the sequel, let us consider a somewhat particular case where the interpersonal relations are prescribed by certain social norms. Menger(1974) considered the clustering problem of such groups into cliques based on a set-theoretic operation on the set of norms.

In general, normative concepts, when considered in the logical framework, can be described in terms of the language in *modal logic* (cf. von Wright[1968], Huges & Cresswell[1968], and Katai, Iwai,et al.[1976 a, b, & c]). For instance,suppose that two propositions p and q represent certain social norms, respectively. Then, there will be four possible attitudes s_1, s_2, s_3, and s_4 tword the norms as follows:

$$s_1 = p \wedge q,\ s_2 = p \wedge \sim q,\ s_3 = \sim p \wedge q, \text{ and } s_4 = \sim p \wedge \sim q, \tag{5.4.91}$$

where s_1 represents the attitude of approving (or admitting)both p and q, while s_2 represents the attitude of approving p and disapproving q etc. When a member v_i considers p as an obligatory norm, then v_i's relation to p can be written as $\Box p$ (or, equivalentlently, as $\sim \Diamond \sim p$) which is read as it is obligatory that p is the case. Also, when v_i considers p as forbidden, then it is written as $\sim \Diamond P$ (or, equivalently, as $\Box \sim p$) which is read as it is forbidden that p is the case). The symbols $\Box$ and $\Diamond$ are called *modal operators* ($\Box$ stands for obligation and $\Diamond$ stands for permission). In general, the mode of v_i's relation to a proposition p can be classified into the following three types: $\Box p$, $\Box \sim p$, and $\Diamond p \wedge \Diamond \sim p$ which are read as p is obligatory, forbidden, and indifferent, respectively. In terms of the notation of state set $S(v_i)$, the above three modes are written as $S(v_i) \quad \{p\}$, $S(v_i) \quad \{\sim p\}$, and $S(v_i) = \{p, \sim p\}$, respectively, as shown in Fig.5.4.12. In the case of two norms p and q, there are 15 possible modes.

It seems to be in accordance with the above logical framework for attitudes to assume that the relationship between an arbitrary pair of members v_i

and v_j is prescribed as

$$R_{ij}(s_i,s_j) \quad \begin{cases} 1 & \text{if } s_i \equiv s_j \\ -1 & \text{if } s_i \not\equiv s_j. \end{cases} \tag{5.4.92}$$

Let us consider the group consisting of five members $v_1,v_2,\cdots,v_5$, and two norms p and q as shown in Fig. 5.4.13, where s_i's are given by (5.4.91). Namely, v_1 considers that norms p and q should be incompatible, and the mode can be written as $\Box(p \not\equiv q)$, while v_2 considers that they should be compatible i.e., $\Box(p \equiv q)$ is true. The modes of v_3, v_4 and v_5 can be written as $\Box p$, $\Box q$, and $\Box(\sim p)$, respectively. It is easy to see that the group is partially balanced, and the unique balanced state is attained when v_1,v_2,v_3,v_4, and v_5, take their attitudes as $\sim p \wedge q$, $p \wedge q$, $p \wedge q$, $\sim p \wedge q$, and $\sim p \wedge q$, respectively. Namely, the group is divided into two cliques $\{v_2, v_3\}$ and $\{v_1, v_4, v_5\}$. The norm q is commonly held by all the members, and the norm p divides the group into the cliques. That is to say, q will be regarded as the *common norm* in the group, and the norm p can be regarded as the *dividing norm*. When only one norm such as p or q is taken into account, then it is easy to see from theorems 5.4.9 and 5.4.10, that any group is totally balanced. In order to characterize the aforementioned balance types in this framework, we introduce the next notation:

$$V(s_i) \triangleq \{v_j \ ; \ s_i \in S(v_j)\}. \tag{5.4.93}$$

The next characterization theorem is a direct consequence of the Structure Theorem.

THEOREM 5.4.12 : When $G(\mathcal{R})$ is complete, and R_{ij}'s are prescribed by (5.4.92), then the following statements hold.

$\mathcal{R}$ is totally balanced iff $\#\{i \ ; \ V(s_i) \not\equiv \phi\} \leq 2$. (5.4.94)

$\mathcal{R}$ is totally unbalanced iff $V(s_i) \cup V(s_j) \not\equiv V$

for any pair of s_i and s_j. (5.4.95)

$\mathcal{R}$ is partially balanced iff $\#\{i \ ; \ V(s_i) \not\equiv \phi\} > 2$,

and there exist s_i and s_j such that $V(s_i) \cup V(s_j) \quad V$. (5.4.96)

In the above, V is the set of members in the group $\mathcal{R}$. Therefore, the

characterization of the balance types in this case is reduced to a simple set-theoretic relationship in the collection of $V(s_i)$'s. The theorem says that a totally balanced group is ruled by at most two different norms, and a totally unbalanced group is a group in which no two norms cannot cover all the members. A partially balanced group is ruled by at least three norms such that two of which cover all the members.

5.5 Conclusions

As mentioned in Section 5.2, the notion of structural balance of a group is based on a simplified idea of 2-valued logic such as a friend (enemy) of my friend is my friend(enemy) and a friend (enemy) of my enemy is my enemy (friend) (cf. Abelson & Rosenberg(1958) and Lambert(1965)). This 2-valued structure leads to the tight bipolar balanced situation summarized as the Structure Theorem.

In Section 5.3.1, we have shown that this tight bipolar configuration of a balanced group is reflected in the principal components of the adjacency matrix of the group (cf. theorem 5.3.1). To characterize the above tight bipolar configuration, we introduced the notion of sign vectors in section 5.3.2. It has been shown that centroid factors play the same role for unbalanced group structures as the principal components for balanced group structures. The number of the elements (lines) in a minimum balancing set for an unbalanced group structure can be considered as a measure of degree of unbalance of the group, and hence, the value indicated by (5.3.20) can be regarded as a measure of balance. Therefore, a group having an adjacency matrix with a simple structure in the sense of centroid factor analysis is more balanced than a group with a complicated adjacency matrix.

In Section 5.3.3, introducing the notion of block decomposition, we have shown that the balancing or the minimum balancing sets for a social group can be decomposed to its block components. Also, by introducing the notion of dual graph, the aggregation problems for social groups with planar graph structures are effectively treated. From theorem 5.3.7, the degree of unbalance can be regarded as a degree of dispersion of the negative vertices in the dual graph. In Section 5.3.4, we have shown that our methods are applicable to the groups with directed member-member relations. We also examined a quantitative approach incorporating the notion of relative importance of each relation inside groups. Another kind of balancing sets with actual importance is the minimal balancing sets introduced by Harary. In Section 5.3.5, we have

developed some methods to derive the minimal balancing sets, by extending the methods in Sections 5.3.2 and 5.3.3.

The researches on the properties of social balance, so far, were concerned only with static groups, i.e., given fixed signed graphs. However, in actual social groups, the signs of interpersonal relations are not constant but fluctuate between positive and negative. We discussed the properties of social balance in such cases by a statistical point of view and a finite-state systems theoretic approach in Section 5.4.

In Section 5.4.1, we first examined the effect of graph structure on their balance. Roughly speaking. from equations (5.4.14), (5.4.21) and (5.4.22), the probability of balance of a group depends on the number of cycles (k) in the group. In other words, the probability of balance decreases when the number of relations (m) or block components (t) increases or the number of members(n) decreases (cf. equation (5.4.22')). The number (n-t) is the minimum number of relations that is sufficient to span the members in a group with (t) block components. Therefore, (m+t-n) is the number of surplus relations in the group. That is, the number (m+t-n) can be considered as a measure of complexity of social groups in terms of their graph structures and the probability of balance has a negative dependence with the above measure. In other words, the more a group becomes to be complex, the less the probability of balance will be.

In Section 5.4.2, we introduced the notion of balance types of social groups based on a finite-state systems theoretic approach. We also considered the characterization problem of these balance types, i.e., totally balanced, totally unbalanced and partially balanced. For totally balanced groups, we introduced the idea of partition on the internal states set of each member and stated the necessary and sufficient condition for a group to be totally balanced (cf. theorems 5.4.7, 5.4.9, and 5.4.10). However, for totally unbalanced or partially balanced groups, their characterization is difficult, and we only stated necessary or sufficient condition (cf. theorem 5.4.8, corollary 5.4.3 and theorem 5.4.10).

Briefly speaking, a totally balanced group is ruled by two different opinions (partitions of internal states) such that there exists a complete consensus stated by equations (5.4.70) and (5.4.71) (cf. corollary 5.4.4). If the above consensus is completely disorganized in the sense of equation (5.4.83), then the group is totally unbalanced (cf. theorem 5.4.11).

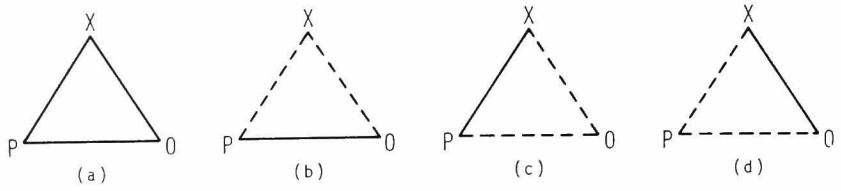


Fig. 5.2.1. Four balanced states in Heider's two-unit one-object group, where P and O are personal units and X is impersonal object. Solid and broken lines correspond to positive and negative relations, respectively.

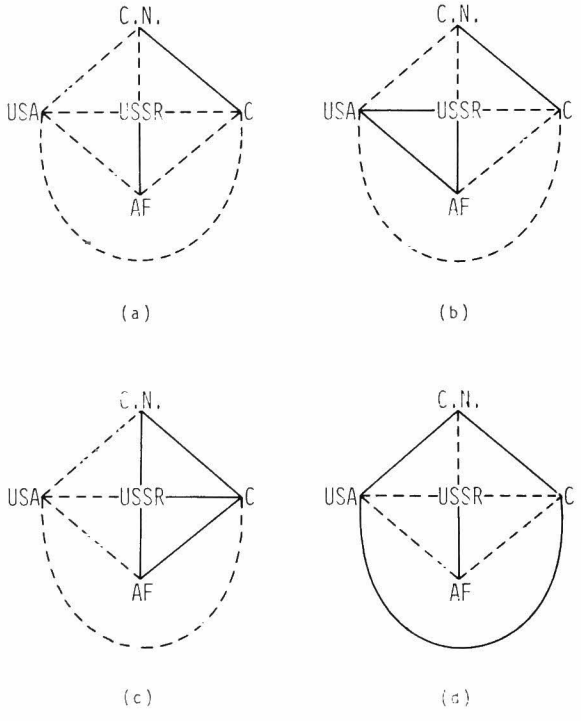


Fig. 5.2.2. Balanced and unbalanced situations of three-nation two-affair group. Situation (a) is unbalanced(the political situation in 1977), and the others are balanced.

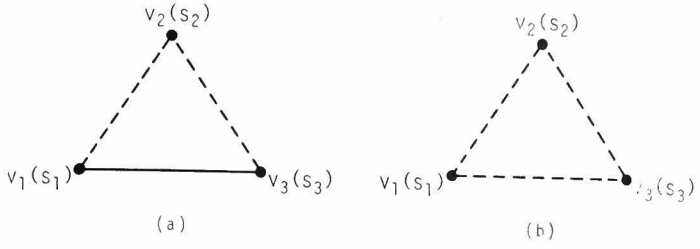


Fig. 5.3.1. Illustrative examples of Lemma 5.3.1.

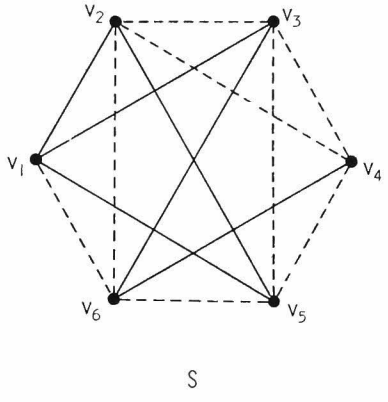


Fig. 5.3.2. An example of unbalanced groups.

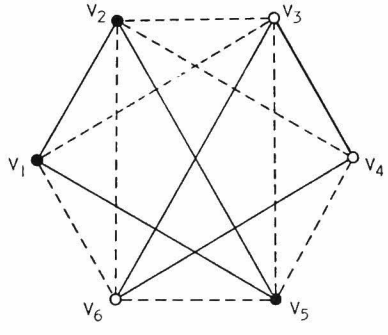


Fig. 5.3.3. An example of balanced groups, which is obtained by changing the relations v_1v_3 and v_3v_4 in the group shown in Fig. 5.3.2.

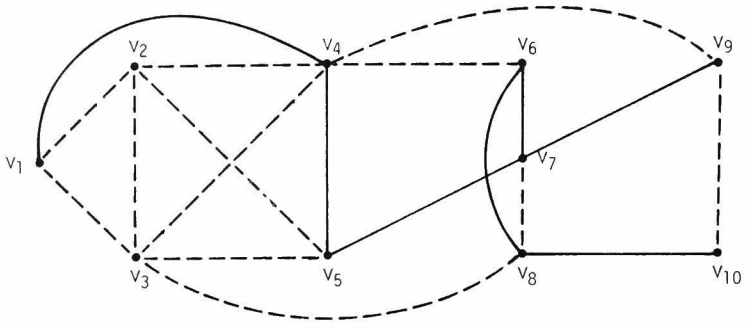


Fig. 5.3.4. An example of unbalanced groups consisted of ten members.

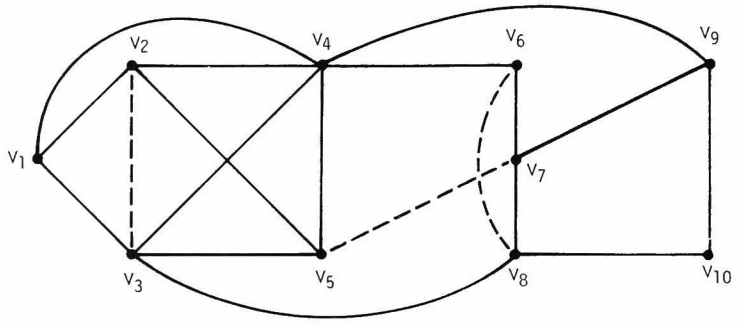


Fig. 5.3.5. A minimum balancing set(which is indicated by broken lines) for the group in Fig. 5.3.4.

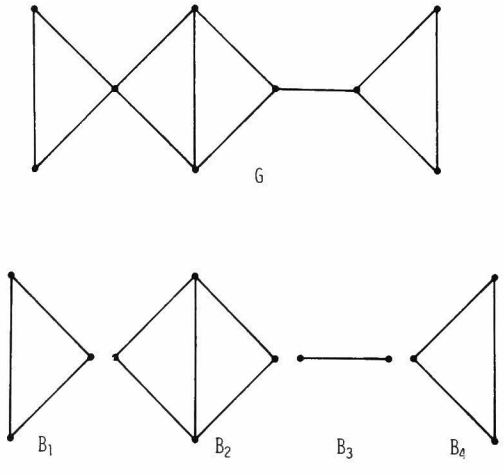


Fig. 5.3.6. Block decomposition of a group G, where subgroups $B_1 - B_4$ are called block components of G.

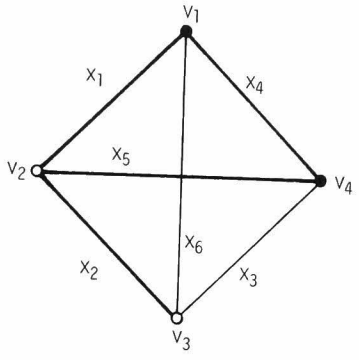


Fig. 5.3.7. Illustrative example of the boudary of a 1-chain.

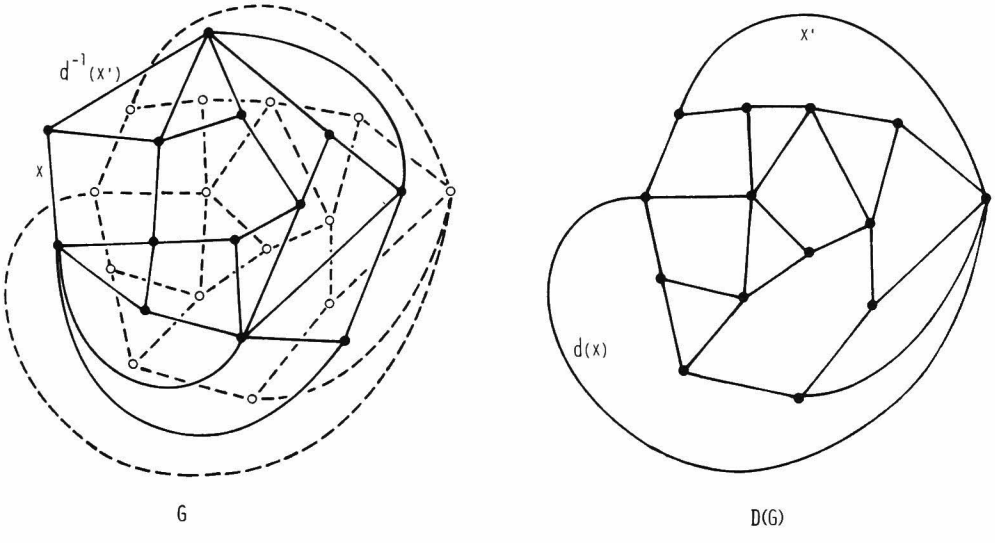


Fig. 5.3.8. Dual graph D(G) of a graph G.

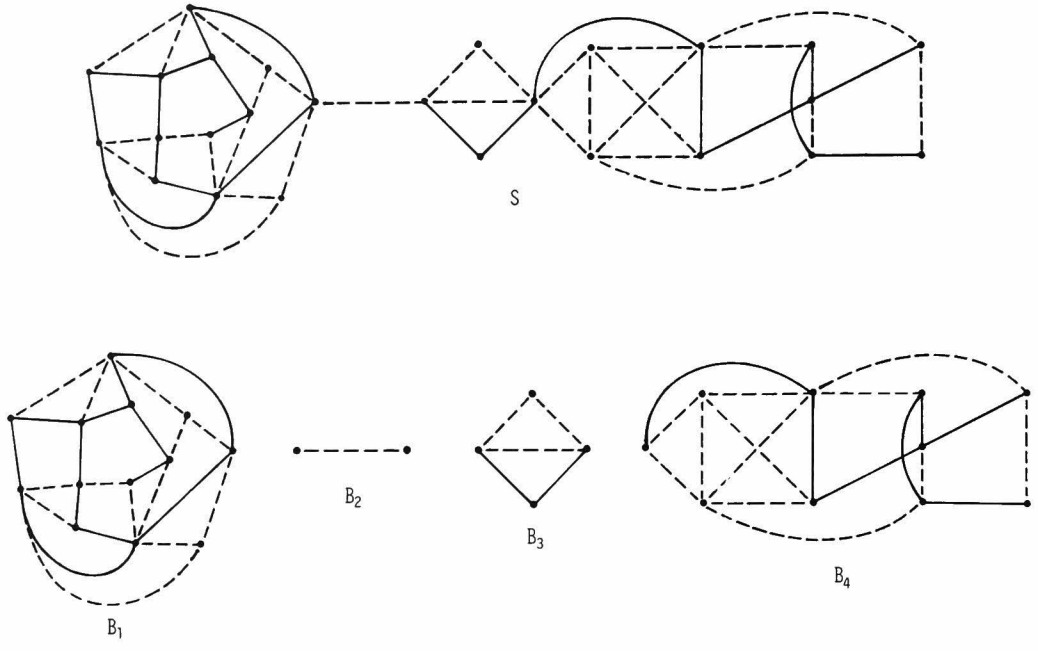


Fig. 5.3.9. Block decomposition of a signed graph S.

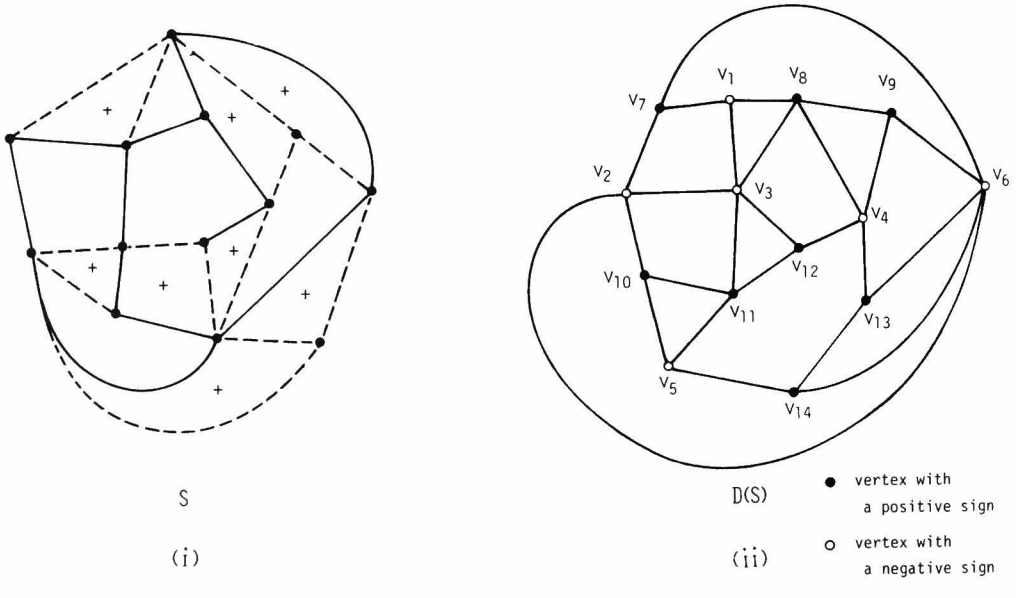


Fig. 5.3.10. A planar signed graph S and its dual graph D(S).

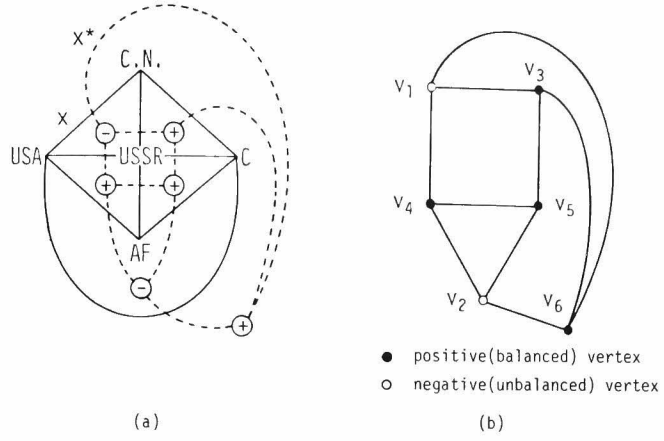


Fig. 5.3.11. Construction of dual graph of three-nation two-affair group under situation in Fig. 5.2.2(a).

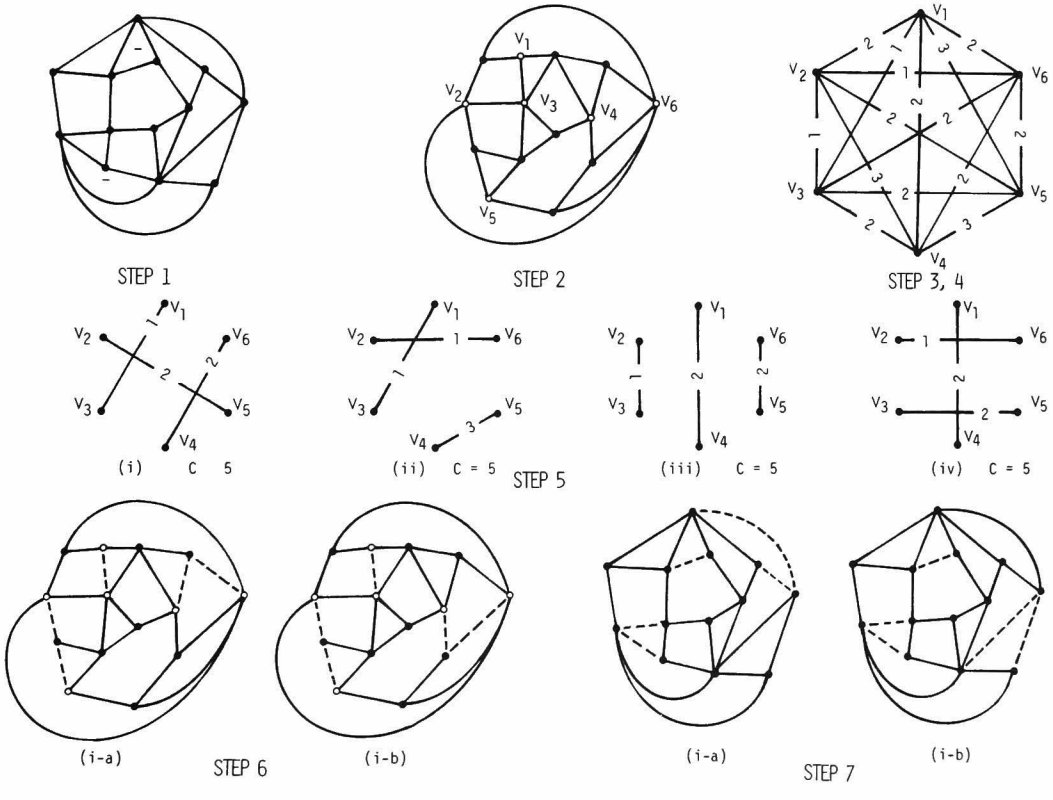


Fig. 5.3.12. Derivation of the minimum balancing sets for the signed graph S in Fig. 5.3.10(i). In this case, there are four optimum pairings (i) - (iv) as shown in step 5. For example, pairing (i) yields two minimum balancing sets (i-a) and (i-b) in step 7, because v_4 and v_6 are joined by two different shortest paths v_4v_9, v_9v_6 and v_4v_{13}, $v_{13}v_6$ as shown in step 6.

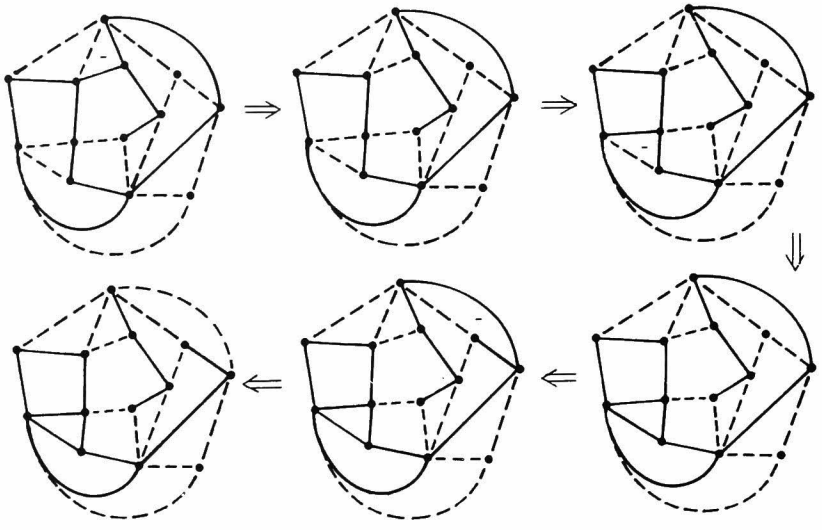


Fig. 5.3.13. The minimum balancing process corresponding to case (i-a) of step 6 in Fig. 5.3.12.

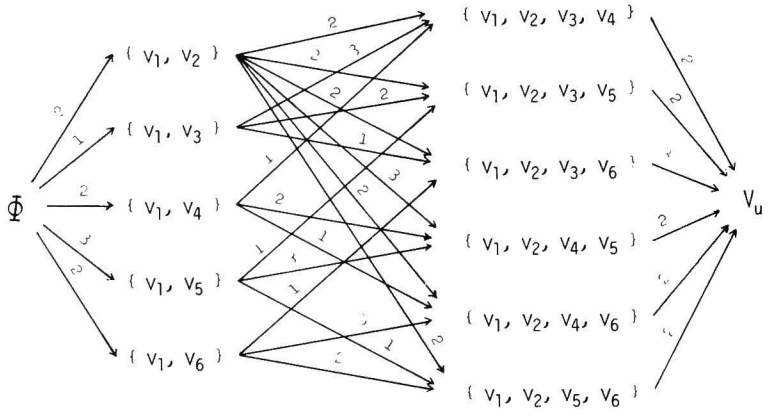


Fig. 5.3.14. Representation of step 5 in Fig. 5.3.12 as a shortest route problem. Each of the shortest routes corresponds to (in a one-to-one manner) an optimum pairing for the group in Fig. 5.3.10(i).

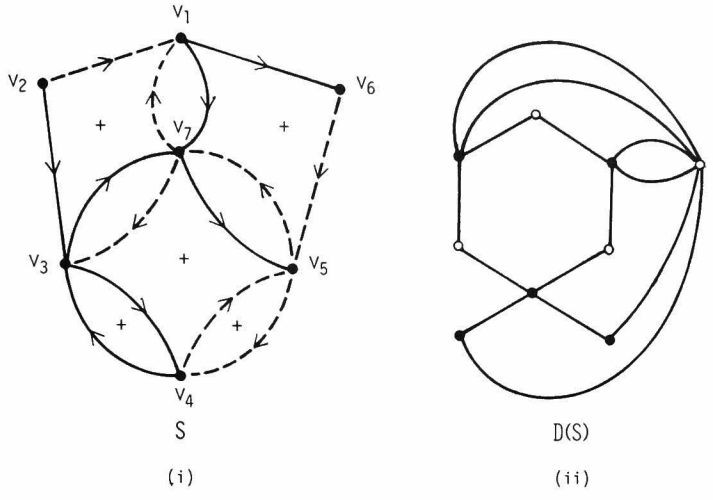


Fig. 5.3.15. Construction of a dual graph D(S) of a planar signed directed graph.

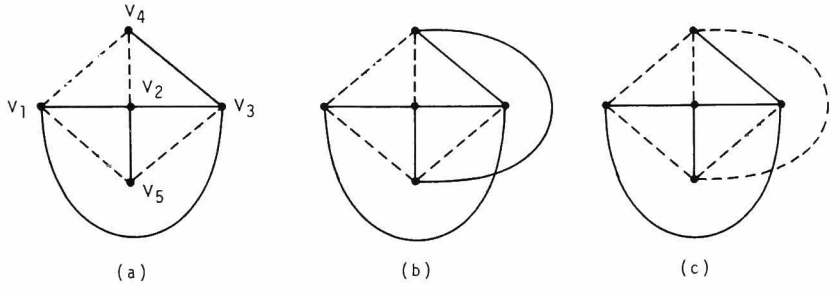


Fig. 5.3.16. Illustrative examples of coalition formation, where v_1, v_2, and v_3 are units and v_4 and v_5 are objects commonly interested by the units. In group (a), there is no relation between the objects. A positive and a negative relation exist in (b) and (c), respectively.

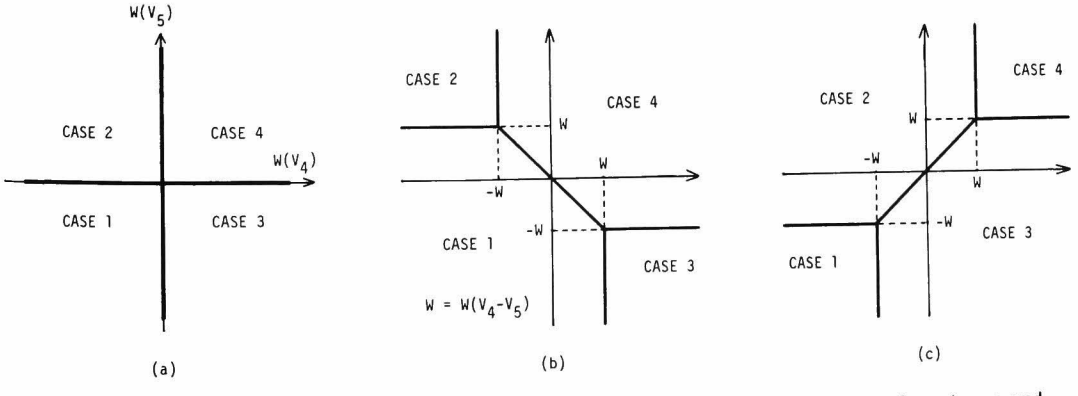


Fig. 5.3.17. The minimum balancing sets for the groups in Fig. 5.3.16 have four types and are determined by indices $w(v_4)$ and $w(v_5)$ given by (5.3.96).

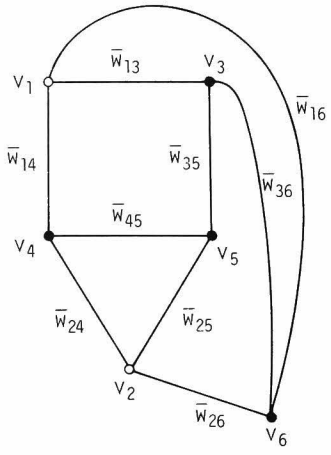
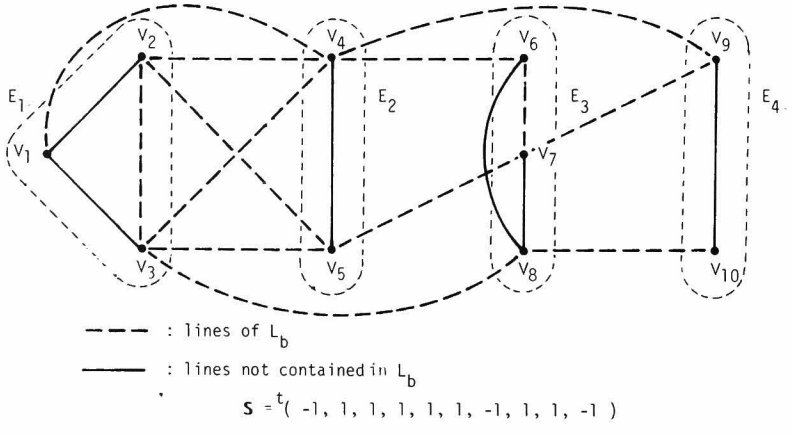
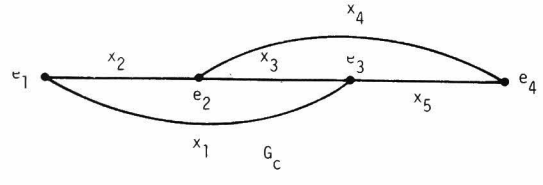


Fig. 5.3.18. Dual graph of the three-nation two-affair group in Fig. 5.2.2(a), where $\bar{w}_{ij}$ represents the relative importance of relation $v_i v_j$.



(i)



(ii)

Fig. 5.3.19. Construction of the condensed graph G_c corresponding to the group S in Fig. 5.3.4 and the balancing set L_b shown by (i).

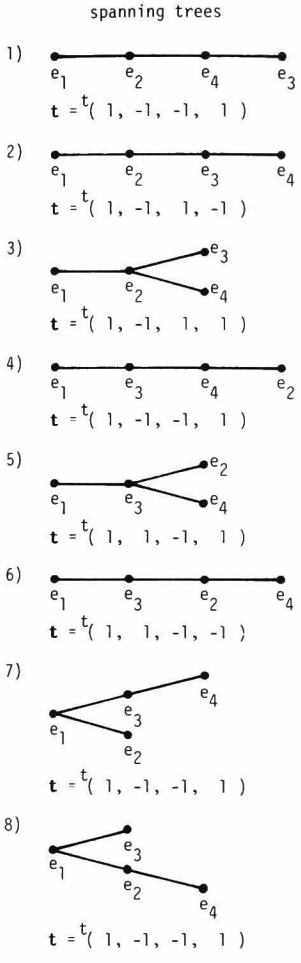


Fig. 5.3.20. The spanning trees of the condensed graph G_c in Fig. 5.3.19(ii).

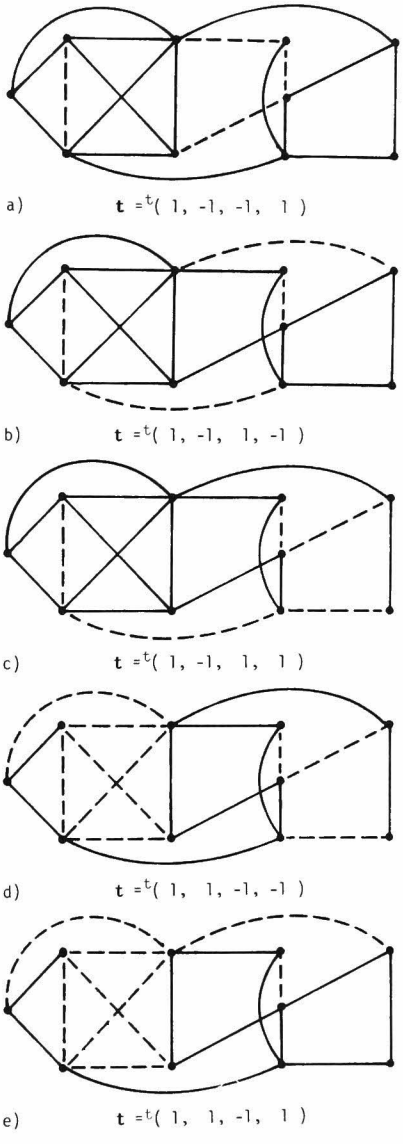


Fig. 5.3.21. The minimal balancing sets in the balancing set shown in Fig. 5.3.19(i).

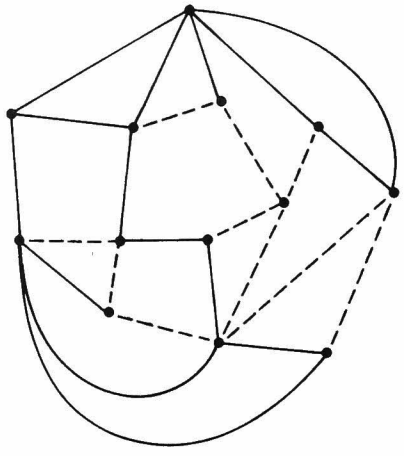


Fig. 5.3.22. A balancing set L_b for the group S in Fig. 5.3.10(i), where the lines of L_b are shown by broken lines.

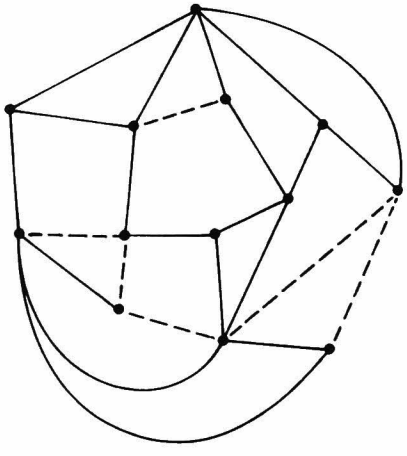
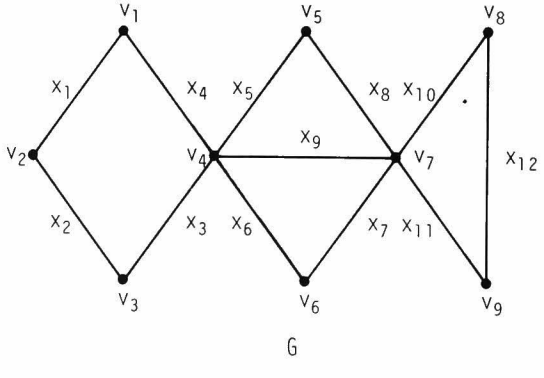


Fig. 5.3.23. A minimal balancing set for the group S. The set is contained in the balancing set in Fig. 5.3.22.



G

Fig. 5.4.1. A graph G composed of 16 cycles and 5 circuits.

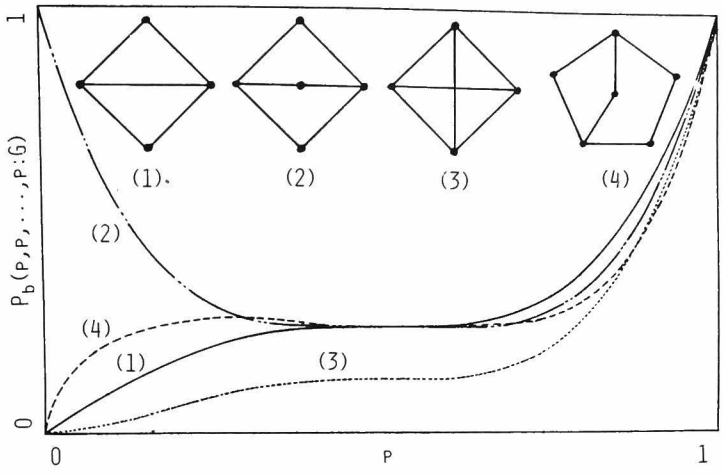


Fig. 5.4.2. Balance probabilities $P_b(p_1,p_2,\cdots,p_n : G)$ of groups (1) - (4) under the condition that $p_1 \quad p_2 \quad \cdots \quad p_n \quad p$.

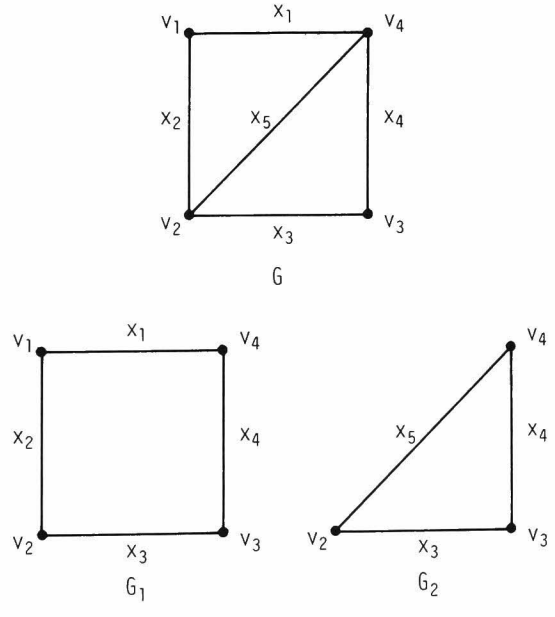


Fig. 5.4.3. Two subgraphs G_1 and G_2 of the graph G negatively correlate to each other when the probabilities $p_1 \quad p_2 \quad \cdots \quad p_5 \quad p$ are less than 0.5.

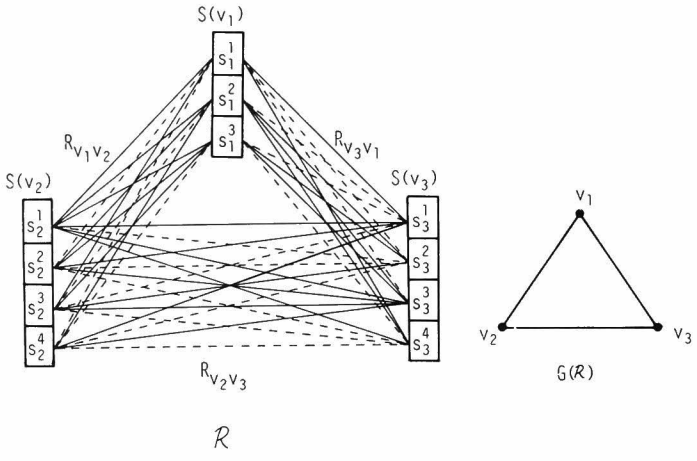


Fig. 5.4.4. A group with a relational structure R composed of three members v_1, v_2, and v_3, where solid and broken lines correspond to positive and negative interpersonal relations, respectively. $G(R)$ represents the graph structure of the group R.

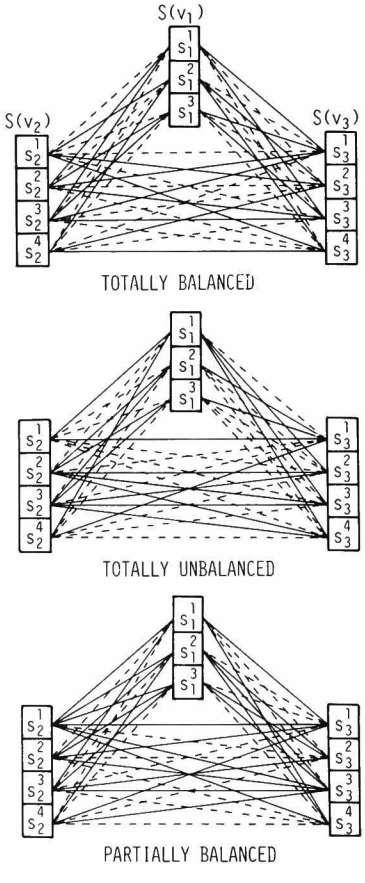


Fig. 5.4.5. Examples of the three balance types composed of three members similar to Fig. 5.4.4.

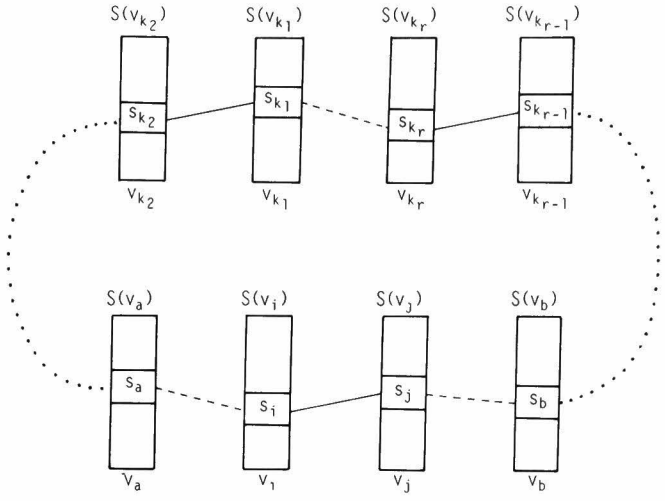


Fig. 5.4.6. An illustration of the method to indicate the partition on the sets $S(v_i)$ and $S(v_j)$ used in the proof of Lemma 5.4.4.

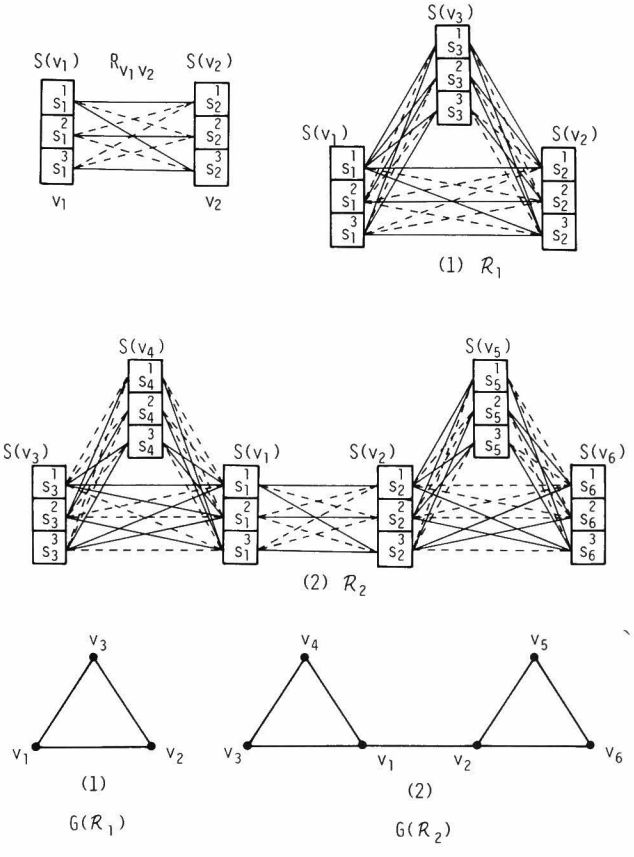


Fig. 5.4.7. A partially balanced group $\mathcal{R}_1$ (shown in (1)) and a totally balanced group $\mathcal{R}_2$ (shown in (2)) containing the members v_1 and v_2 whose interpersonal relation $R_{v_1v_2}$ satisfies the condition of Corollary 5.4.3. The graph $G(\mathcal{R}_1)$ is strictly nonseparable, while $G(\mathcal{R}_2)$ is not.

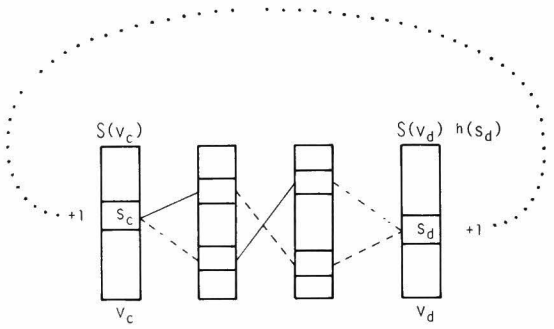


Fig. 5.4.8. An illustrative example of the method in the proof of theorem 5.4.9 for the sign assignment $h(s_d)$ to an internal state s_d in $S(v_d)$ based on the sign +1 of the internal state s_c in $S(v_c)$.

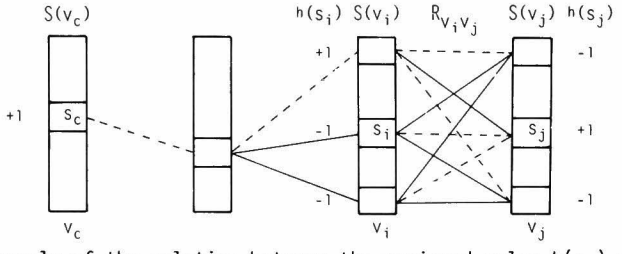


Fig. 5.4.9. An example of the relation between the assigned value $h(s_i)$ and $h(s_j)$ to s_i in $S(v_i)$ and s_j in $S(v_j)$ such that v_i and v_j are adjacent(cf. the proof of Theorem 5.4.9).

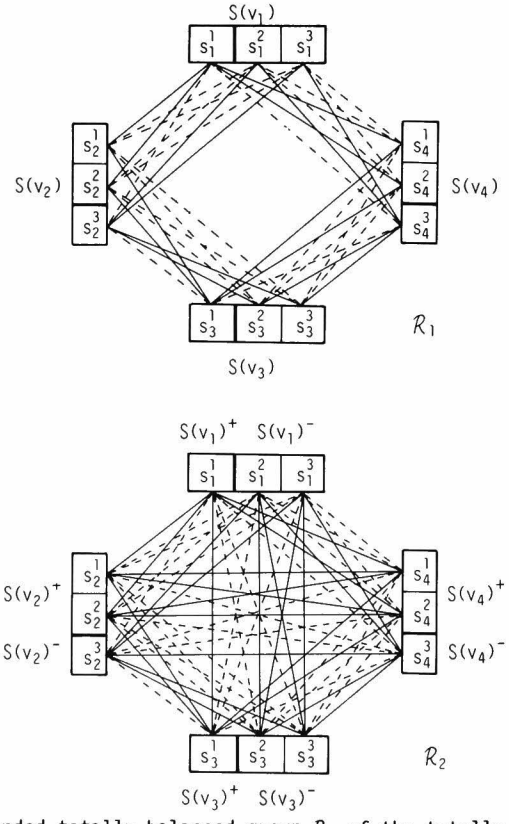
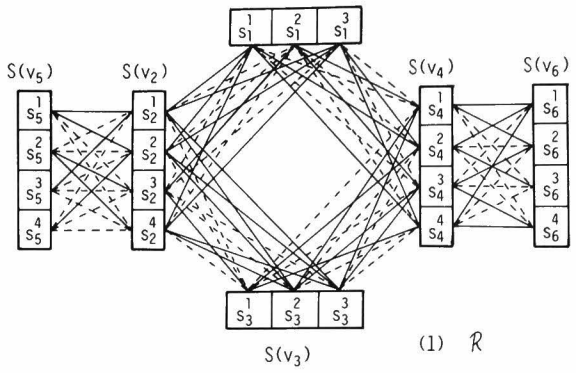


Fig. 5.4.10. The extended totally balanced group R_2 of the totally balanced group R_1 such that the graph structure $G(R_2)$ of R_2 is complete while $G(R_1)$ is an incomplete graph.



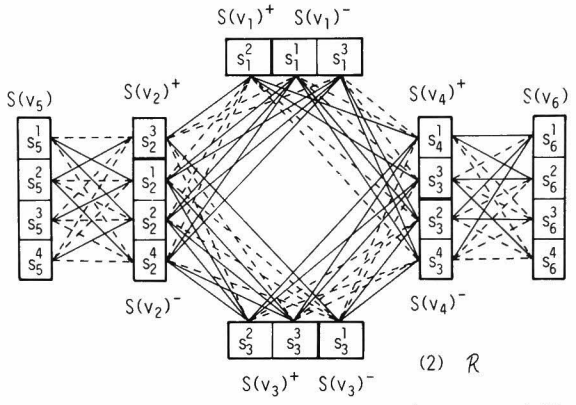


Fig. 5.4.11. An example of a totally balanced group R(shown in (1)) such that its graph structure G(R) contains just one circuit composed of four members v_1, v_2, v_3 and v_4. From Theorem 5.4.11, the internal state sets $S(v_1)$, $S(v_2)$, $S(V_3)$, and $S(v_4)$ can be partitioned as illustrated in (2), where k_i i for i 1, 2, 3, 4 and q 2(cf. equations (5.4.82) and (5.4.83)).

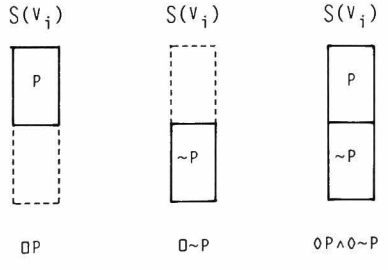


Fig. 5.4.12. State set $S(v_i)$ and modal relation to a norm p of member v_i.

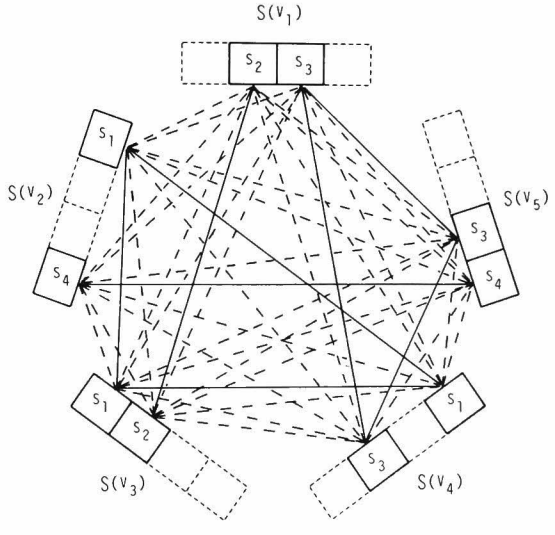


Fig. 5.4.13. A partially balanced group which is consisted of five members and prescribed by two norms p and q(cf. (5.4.91)

	s^0 d_j	s^1 d_j	s^2 d_j	s^3 d_j	s^0 d_j	s^1 d_j	s^2 d_j	s^0 d_j	s^1	s^0 d_j	s^1	s^0 d_j	s^1 d_j	s^2 d_j	s^0 d_j	s^1 d_j	s^2	s^0 d_j	s^1 d_j	s^2 d_j	s^3 d_j	s^4 d_j	···
v_1	1 -1	1 1	1 3	1 3	-1 1	-1 3	-1 3	1 1	1	1 1	1	1 -3*	1 (-1)	-1 1	1 -1	1 (-3)	-1	1 -1	1 1	1 3	1 3	1 3	···
v_2	1 -4	1 (-2)	-1 2	-1 2	1 -2	1 0	1 2	-1 4	-1	1 (-2)	-1	1 -2	1 0	1 2	1 -2	1 0	1	1 -4	1 (-2)	-1 2	-1 2	-1 2	···
v_3	1 (-5)	-1 5	-1 3	-1 3	1 -3	1 -1	1 1	1 (-3)	-1	-1 5	-1	1 (-3)	-1 3	-1 1	1 -3	1 -1	1	1 (-5)	-1 5	-1 3	-1 3	-1 3	···
v_4	1 -2	1 0	1 2	1 4	1 (-4)	-1 4	-1 6	1 0	1	1 0	1	-1 2	-1 0	-1 2	1 (-4)	-1 4	-1	1 0	1 2	1 4	1 6	1 6	···
v_5	1 0	1 2	1 4	1 4	1 0	1 (-2)	-1 2	1 2	1	1 2	1	1 -2	1 0	1 0	-1 0	-1 2	-1	1 0	1 2	1 4	1 4	1 2	···
v_6	1 1	1 1	1 1	1 1	1 1	1 3	1 3	1 1	1	1 1	1	1 3	1 3	1 3	1 1	1 3	1	-1 -1	-1 -1	-1 -1*	-1 -1	-1 1	···
v_7	1 2	1 2	1 2	1 0	1 2	1 2	1 0	1 2	1	1 2	1	1 2	1 2	1 2	1 0	1 0	1	1 0	1 0	1 0	1 (-2)	-1 2	···
v_8	1 0	1 2	1 2	1 2	1 0	1 0	1 0	1 0	1	1 2	1	1 0	1 2	1 2	1 0	1 0	1	1 -2	1 0	1 0	1 0	1 2	···
v_9	1 -1	1 -1	1 (-1)	-1 1	1 -1	1 1	1 1	1 -1	1	1 -1	1	1 1	1 1	1 1	1 -1	1 1	1	1 -1	1 -1	1 (-1)	-1 1	-1 3	···
v_{10}	1 0	1 0	1 0	1 2	1 0	1 0	1 0	1 0	1	1 0	1	1 0	1 0	1 0	1 0	1 0	1	1 0	1 0	1 0	1 2	1 2	···
$\hat{d}(\mathbf{s}^{\cdot})$	-10	10	18	22	-6	10	18	6		10		-2	10	14	-10	6		-14	6	14	18	26	···
$c(L_b)$	12	7	5	4	11	7	5	8		7		10	7	6	12	8		13	8	6	5	3	···

Table 5.3.1. An illustration of the algorithm to derive the minimal balancing sets for the group in Fig. 5.3.4. The encircled numerals represent the minimum value of $d_j(\mathbf{s})(j = 1, 2, \cdots, 10)$ at each step(column); the numerals with * also mean the minimum values which are not used to designate the position α_i indicated by (5.3.41) to avoid duplications of the procedure; i.e., if they are used to designate α_i, the resultant sign vectors are the same as those which appeared in previous steps.

	X_1	X_2	X_3	X_4	X_5
1)	1	0	1	0	0
2)	1	0	0	1	0
3)	1	0	0	0	1
4)	0	1	1	0	0
5)	0	1	0	1	0
6)	0	1	0	0	1
7)	0	0	1	1	0
8)	0	0	1	0	1

Table 5.3.2. List of the lines deleted from the condensed graph G_c in Fig. 5.3.19(ii). Each of cases 1 - 8 yields a spanning tree of G_c as shown in Fig. 5.3.20.

REFERENCES

Abelson, R. P. & Rosenberg, M. J.
(1958) Symbolic psycho-logic: a model of attitudinal cognition. Behavioral Science, vol. 3, pp. 1 13.

Abelson, R. P. & Tukey, J. W.
(1963) Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order, Annals of Mathematical Statistics, vol. 34, pp. 1347 1369.

Anderson T. W.
(1958) An Introduction to Multivariate Analysis. John Wiley.

Aoyama, H.
(1965a) Dummy variable and its application to the quantification methods, Proceedings of the Institute of Statistical Mathematics, vol. 13, No.1, pp. 1 12,(in Japanese).
(1965b) Erata on "Dummy variable and its application to the quantification method", Proc. Inst. Statist. Math. Vol. 13, No. 2, pp. 135 137, (in Japanese)

Ara, K.
(1959) The aggregation problem in input - output analysis, Econometrika, Vol. 27, pp. 257 262.

Babu, C. C.
(1972a) On the application of divergence for the extraction of features from imperfectly labeled patterns, IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC - 2, No. 2, pp. 290 - 292.
(1972b) On the distance criterion of Patric and Fischer, IEEE Trans. on Information Theory, Vol.IT - 18, No. 3, p. 428.

Barndorff - Nielsen, O.
(1974) Subfields and loss of information,Z. Wahrsheinlichkeitstheorie, Vol. 2,pp. 369 - 379.

Barr, D.R. & Thomas, M. U.
(1977) An Eignvector Condition for Markov Chain Lumpability, Operations Research, Vol. 25, No. 6, pp. 1028 1031.

Baxa, E. G. Jr. & Nolte, L. W.
(1972) Adaptive signal detection with finite memory. IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC - 2, No. 1, pp. 42 - 49.

Bell, C. B.
(1958) On the structure of stochastic independence, Illinois Journal of Mathematics,Vol.2, pp 415 - 424.
(1962) Mutual Information and Maximal Correlation as Measures of Dependence, Annals of Mathematical Statistics, Vol. 33, No. 2, pp. 587 595.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W.
(1975) Discrete Multivariate Analysis : Theory and Practice, MIT Press, Mass.

Bourton, D. M. & Wallace, C. S.
(1969) The information content of a multistate distribution, Journal of Theoretical Biology, Vol. 23, pp. 269 - 278.
(1970) A program for numerical classification, The Computer Journal Vol. 13, No. 1, pp. 63 - 69.

Burke, C. J. & Rosenblatt, M.
(1958) A Markovian Function of a Markov Chain, Annals of Mathematical Statistics, Vol. 29, pp. 112 122.

Bush, R. R. & Galanter, E. & Luce, R. D.
(1959) Tests of the "Beta Model", In R. R. Bush & W. K. Estes (Eds.), Studies in Mathematical Learning Theory, Stanford: Stanford Univ. Press.

Bush, R. R. & Mosteller, F.
(1955) Stochastic Models for Learning, Wiley.

Caprihan, A. & De Figneiredo, R. J. P.
(1970) On the extraction of pattern features from continuous measurements, IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC - 6, No. 2, pp. 110 - 115.

Carlton, A. G.
(1969) On the bias of information estimates, Psychological Bulletin, Vol. 71, No. 2, pp. 108 - 109.

Carroll, M. P.
(1973) Applying Heider's Theory of Cognitive Balance to Claude Lévi-strauss, Sociometry, Vol. 36, pp. 285 - 301.

Cartwright, D. (Ed.)
(1959) Studies in Social Power, Ann Arbor, Michigan Univ. Press.

Cartwright, D. & Harary, F.
(1956) Structural balance: a generalization of Heider's theory, Psychological Review, Vol. 63, No. 5, pp. 277 293.

Chernoff, H.
(1959) Sequential design of experiments, Annals of Mathematical Statistics, Vol. 30, pp. 755 770.

Chino, S.
(1963) The relation between correlation ratio and success rate in the classification by quantification method, Proceedings of the Institute of Statistical Mathematics, Vol. 11, No. 1, pp. 7 24, (in Japanese).

Chow, C. K.
(1966) An approach to structure adaptation in pattern recognition, IEEE Trans. on Systems Science and Cybernetics, Vol. SSC 2, No. 2, pp. 73 - 80.

Chow, C. K. & Liu, C. N.
(1968) Approximating discrete probability distributions with dependent trees, IEEE Trans. on Information Theory, Vol. IT - 14, No. 3 . pp. 462 467.

Cover, T. M.
(1972) A Hierarchy of probability density function estimates, pp. 83 - 98, in S. Watanabe (Ed.) Frontiers of Pattern Recognition, Academic, Press.

Csiszár, I.
(1975) I divergence geometry of probability distributions and minimization problems, Annals of Mathematical Statistics, Vol.3, No. 1, pp. 146 158.

Daróczy, Z.
(1970) Generalized information functions, Information and Control, Vol. 16, pp 36 51.

Davidsen, D., Suppes, P., & Siegel, S.
(1957) Decision Making: an exp. approach, Stanford: Stanford Univ. Press.

Davis, J. A.
(1967) Clustering and structural balance in graphs, Human Relations, Vol. 20, pp. 181 - 187.

Davis, J. D.
(1968) "Statistical analysis of pair relationships: symmetry, subjective consistency and reciprocity," Sociometry, Vol. 31, pp. 102 119.

Doob, J. L.
(1953) Stochastic Processes, New York: Wiley, pp. 292 - 318.

Dubins, L. E.
(1962) Rises and upcrossings of nonnegative martingales, Illinois Journal of Mathematics, Vol. 6, pp. 226 - 241.

Econometric Committee in Economic Deliberation Council of Japan
(1970) The third report from the Econometric Comittee: Report on the econometric model for the new economic developmental planning, (in Japanese).

Endo, J. & Iwai, S.
(1974) Medical information contained in chemical graph patterns measured by SMA 12 / 60, Proceedings of Technikon International Symposia on "Survival and Society", Techonikon Japan, Tokyo, pp. 73 - 81.

Feather, N. T.
(1965) A structural balance analysis of evaluative behavior, Human Relations, Vol. 18, pp. 171 185.

Festinger, L.
(1949) The analysis of sociograms using matrix algebra, Human Relations, Vol. 2, pp. 153 158.

Fisher, R. A.
(1959) Statistical Methods and Scientific Inference, Oliver & Boyd Ltd. Edinburgh.

Flament, C.
(1963) Applications of Graph Theory to Group Structure, Englewood Cliffs, NJ.: Prentice-Hall, pp. 89 126.

Foley, D. H.
(1972) Considerations of sample and feature size,IEEE Trans. on Information Theory, Vol. IT - 18, No. 5, pp. 618 - 626.

French, D. Jr.
(1956) A formal theory of social power, Psychological Review, Vol. 63, No. 3, pp. 181 194.

Friedman, H. P. & Rubin, J.
(1967) On some invariant criteria for grouping data, American Statistical Association Journal, Vol.62, No. 320, pp. 1159 - 1178.

Fukao, T.
(1971) On the system identification, decomposition, and consolidation, The Transactions of the Institute of Electronics and Communication Engineers of Japan, Vol. 54 - c, No. 12, pp. 1141 1148.
(1972) On the system decomposition and consolidation problems, Systems and Control, Vol. 16, No. 1, pp. 12 18.

Galanter, E. & Bush, R. R.
(1959) Some T - maze experiments, In R. R. Bush & W. K. Estes (Eds.), Studies in Mathematical Learning Theory, Stanford: Stanford Univ. Press.

Gabelein, H
(1941) Das statistishe Problem der Korrelation als Variations und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung, Zeitshrift fur Angewandte Mathematik und Mechanik, Vol. 21, pp. 364 379.

Ghurye, S. G.
(1968) Information and sufficient sub-fields, Annals of Mathematical Statistics, Vol. 38, No. 6, pp. 2056 2066.

Gish, H. & Pierce, J. N.
(1968) Asymptotically efficient quantizing, IEEE Trans. on Information Theory, Vol. IT - 14, No. 5, pp. 676 - 683.

Good, I. J.
(1963) Maximum entropy for hypothesis formulation, especially for multi-dimensional contingency tables, Annals of Mathematical Statistics, Vol. 34, No. 3, pp. 911 - 934.

Gower, J. C.
(1966) Some distance properties of latent root and vector methods used in multivariate analysis, Biometrika, Vol. 53, pp. 325 338.

Green, B. F.
(1951) A general solution for the latent class model of latent structure analysis, Psychometrika, Vol. 16, pp. 151 156.

Guttman, L.
(1950) The basis for scalogram analysis, In Stouffer et al.(Eds.), Measurement and Prediction,Princeton Univ. Press, N.J.
(1968) A general nonmetric technique for finding the smallest coordinate space for a configuration of points, Psychometrika, Vol. 33, No. 4, pp. 469 - 506.

Hájek, J.
(1969) A Course in Nonparametric Statistics, Holden-day, Inc..

Hamdan, M. A. & Tsokos, C. P.
(1971) An Information Measure of Association in Contingency, Tables, Information and Control, Vol. 19, pp. 174 - 179.

Hansen, J. C.
(1975) Some applications of a general theory of digraph measures, Proceedings of the 1975 International Symposium on Multiple-Valued Logic, Indiana Univ.

Harary, F.
(1955) On local balance and N-balance in signed graph, Mich. Math. J., Vol. 63, pp. 37 - 41.
(1956) On the measurement of structural balance, Behavioral Science, Vol. 4, pp. 316 - 323.
(1959) Graph theoretic methods in the management sciences, Management Science, Vol. 5, pp. 387 - 403.
(1961) A structural analysis of the situation in the Middle East in 1956, Journal of Conflict Resolution, Vol. 5, NO. 2, pp. 167 178.
(1969) Graph Theory, Addison-Wesley.

Harary, F. & Norman, R.Z.
(1953) Graph Theory as a Mathematical Model in Social Science, Ann Arbor, Mich.: Institute for Social Research.

Harary, F., Norman, R. Z. & Cartwright, D.
(1965) Structural Models: an introduction to the theory of directed graphs, Wiley, New York, pp. 339 - 361.

Hartmanis, J.
(1960) Symbolic analysis of a decomposition of information processing machines, Information and Control, Vol. 3, pp. 154 178.

Hartmanis, J. & Stearns, R. E.
(1966) Algebraic Structure Theory of Sequential Machines, Prentice-Hall.

Hatanaka, M.
(1952) Note on Consolidation within a Leontief System, Econometrika, Vol. 20, pp. 301 303.

Hayashi, C.
(1952) On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view, Annals of the Institute of Statistical Mathematics, Vol. 3, No. 2, pp. 69 - 98.

Hayashi, C. & Murayama, T.
(1969) Planning and Practice in Market Researches (Shijō chōsa no keikaku to jissai),Nikkan Kōgyo News Paper Comp., Tokyo, (in Japanese).

Heider, F.
(1946) Attitudes and cognitive organization, Journal of Psychology, Vol. 21, , , , pp. 107 112.

Hellman, M. E. & Cover, T. M.
(1970) Learning with finite memory, Annals of Mathematical Statistics, Vol. 41, No. 3, pp. 765 - 782.
(1971) On memory saved by randomization, Annals of Mathematical Statistics, Vol.42, No. 3, pp. 1075 1078.

Hellman, M. E. & Raviv, J.
(1970) Probability of error, equivocation , and the Chernoff bound, IEEE Trans. on Information Theory, Vol. IT - 16, No. 4, pp. 368 - 372.

Hendrickson, A. D. & Buehler, R. J.
(1971) Proper scores for probability forecasters, Annals of Mathematical Statistics, Vol. 42, No. 6, pp. 1916 1921.

Hilgard, E. R. & Bower, G. H.
(1966) Theories of Learning, Appelton.

Ho, Y. C. & Agrawara, A. K.
(1968) On pattern classification algorithms-introduction and survey, IEEE Trans. on Automatic Control, Vol. AC - 13, No. 6, pp. 676 - 690.

Horst, P.
(1965) Factor Analysis of Data Matrices, Holt, Rinehart and Winston, Inc., pp. 114 136.

Hotelling, H.
(1936) Relations between two sets of variates, Biometrika, Vol. 28, pp. 321 377.

Huges, G. E. & Cresswell, M. J.
(1968) An Introduction to Modal Logic, Methuen.

IBM
(1976) Compa: Component Analysis, IBM Japan Data Center Services, Tokyo, (in Japanese).

Iwahori, N.
(1974) Graphs and Probability Matrices (Gurafu to Kakuritsu gyōretsu), Sangyo Tosho, (in Japanese).

Iwai, S.
(1977) Medical information processing and pattern recognition, Systems and Control, Vol. 21, No. 5, pp. 267 275, (in Japanese).

Iwai, S. & Katai, O.
(1978) Graph theoretic models of social group structures and indices of group structures, Systems and Control, Vol. 22, No. 12, pp. 713 - 722, (in Japanese).

Iwamura, T.
(1966) Lattice Theory (Sokuron), Kyoritzu Zensho 161, Kyoritzu Pub. Comp., (in Japanese).

Jeffreys, H.
(1946) An invariant form for the prior probability in estimation problems, Proc. Roy. Soc. London, Ser. A, Vol. 186, pp. 453 461.

Johnson, P. O.
(1950) The quantification of qualitative data in discriminant analysis, American Statistical Association Journal, Vol. 45, pp. 65 76.

Kailath, T.
(1967) The divergence and Bhattacharyya distance measures in signal selection, IEEE Trans. on Computers, Vol. COM - 15, No. 1, pp 52 - 60.

Kammler, H.
(1974) "Formal models vs. real models: tools of inquiry in international relations," in A. Rapoport (ed), General Systems; Yearbook of the Society for General Systems Research, Vol. 19, Washingron, D. C.: Society for General Systems Research, pp. 135 146.

Kanal, L & Chandrasekaran, B.
(1968) On dimensionality and sample size in statistical pattern classification , Proc. of National Electronics Conf.,Chicago, pp. 2 - 7.

Kaneko, Y.
(1976) Inter - industrial Analysis (Sangyo renkan bunseki), Yuuhikaku, (in Japanese)

Kaplan, M. A.
(1967) System and Process in International Politics. New York: Wiley.

Karlin, S.
(1966) A First Course in Stochastic Process,Academic Press, N. Y.

Katai, O., Imanaga, T. & Iwai, S.
(1972a) A consideration on probability theoretical structure of pattern space and its optimum quantization problem, Preprints of the SICE (Society of Instrument and Control Engineers) symposium, pp. 121 122, Tokyo, (in Japanese).
(1972b) Information theoretical consideration on optimum partition and quantization of probabilistic pattern space, Preprints of the 15th Joint Convention on Automatic Control held by JAACE (Japan Association of Automatic Control Engineers), SICE et al., pp. 341 342, Kobe, (in Japanese).

(1974) An informational approach to pattern quantization and a study of its asymptotic properties, Transactions of the Society of Instrument and Control Engineers of Japan, Vol. 10, No. 2, pp. 153 - 159, (in Japanese).
(1975) Analysis and aggregation of inter-industrial relationships based on quantification theory, Preprints of the 18th Joint Convention on Automatic Control held by JAACE, SICE et al., pp. 175 176, Tokyo, (in Japanese).

Katai, O. & Iwai, S.
(1971) Information theoretical considerations on Bayesian type recognition systems, the 14th Joint Convention on Automatic Control held by JAACE, SICE et al., pp. 61 - 62, Tokyo, (in Japanese).
(1972) An information theoretic consideration for Bayesian pattern recognition, Transactions of the Society of Instrument and Control Engineers of Japan, Vol. 8, No. 6, pp. 732 739, (in Japanese).
(1974) Probability theoretical approach for compression of discrimination information, Transactions of the Society of Instrument and Control Engineers of Japan, Vol. 10, No. 1, pp. 86 - 92, (in Japanese).
(1978a) On the characterization of balancing processes of social systems and the derivation of the minimal balancing processes, IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC - 8, No. 5, pp. 337 - 348.
(1978b) Characterization of social balance by statistical and finite-state systems theoretical analysis, Proceedings of the International Conference on Cybernetics and Society held by IEEE, pp. 312 - 316, Tokyo.
(1978c) Studies on the balancing, the minimal balancing, and the minimum balancing processes for social groups with planar and nonplanar graph structures, Journal of Mathematical Psychology, Vol. 18, No. 2, pp. 140 - 176.

Katai, O., Iwai, S., Takagi, S., Yonemoto, K., Tamaru, T., Endo, J. & Fujii, M.
(1973) Introduction of multivariate analysis techniques into blood chemical component patterns, Preprints of the 16th Joint Convention on Automatic Control held by JAACE, SICE et al., pp. 315 - 316, Tokyo, (in Japanese).

Katai, O., Iwai, S., Toyofuku, Y. & Takagi, S.
(1974) Statistical analysis of disease structures reflected on blood chemical component patterns, Preprints of the 17th Joint Convention on Automatic Control held by JAACE, SICE et al., pp. 321 322, Kyoto, (in Japanese).

Katai, O., Iwai, S. & Tamaru, T.
(1975a) Considerations on the balance and dynamics of human group systems, Preprints of the 1st Systems Symposium held by SICE, pp. 195 200, Tokyo, (in Japanese).

Katai, O., Iwai, S., Tamaru, T. & Moribayashi, S.
(1975b) Considerations on the balance and dynamics of social groups, Preprints of the 18th Joint Convention on Automatic Control held by JAACE, SICE et al., pp. 181 182, Tokyo, (in Japanese).

Katai, O., Iwai, S., Nakao, K., Toyofuku, Y. & Moribayashi, S.
(1975c) Probability theoretical considerations on cell proliferating cycles and control of proliferating cycles by the doses of carcinostatic substance, Preprints of the 7th Stochastic Systems Symposium held by JAACE, pp. 109 112, kyoto, (in Japanese).

Katai, O., Iwai, S., Moribayashi, S., Ōtsubo, T. & Iwasaki, T.
(1976a) A consideration on modal logic and logical structure of information processing by social groups, Preprints of the 2nd Systems Symposium held by SICE, pp. 73 78, Kyoto, (in Japanese).
(1976b) A consideration on modal logic, Preprints of the 19th Joint Convention on Automatic Control held by JAACE, SICE et al., pp. 79 80, Osaka, (in Japanese).
(1976c) Logical structural model of information processing process by social groups, Preprints of the 19th Joint Convention on Automatic Control held by JAACE, SICE et al., pp. 191 192, Osaka, (in Japanese).

Katai, O., Iwai, S., Moribayashi, S. & Tamaru, T.
(1976d) A mathematical study on the balance of social groups, Preprints of the SICE Symposium on Mathematical System Theory, pp. 23 - 28, Nagoya.

Kemeny, J. G. & Snell, J. L.
(1963) Finite Markov Chains, D. Van. Norstand Company, Inc., Princeton, N. J.

Kramer, H. P.
(1960) Maximal correlation and uncertainty, Notices Amer. Math. Soc. Vol. 7, Issue No. 50, No. 7, pp. 977 - 978.

Kruskal, J. B.
(1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika, Vol. 29, No. 1, pp. 1 27.
(1964b) Nonmetric multidimensional scaling: a numerical method, Psychometrika, Vol. 29, No. 2, pp. 115 129.

Ku, H. H. & Kullback, S.
(1968) Interaction in multidimensional contingency tables: an information theoretic approach, J.of Research of the National Bureau of Standards - Mathematical Sciences, Vol. 72B, No. 3, pp. 159 199.
(1969) Approximating discrete probability distributions, IEEE Trans. on Information Theory, Vol. IT - 15, No. 4 pp. 444 - 447.

Kullback, S
(1952) Information Theory and Statistics, J. Wiley.
(1968) Probability densities with given marginals, Annals of Mathematical Statistics, Vol. 39, No. 4, pp. 1236 1243.

Kullback, S. & Leibler, R. A.
(1951) On information and sufficiency, Annals of Mathematical Statistics, Vol. 22, pp. 79 86.

Lambert, R. M.
(1966) An examination of the consistency characteristics of Abelson and Rosenberg's "Symbolic psycho-logic", Behavioral Science, Vol. 11, No. 2, pp. 126 - 130.

Laming, D. R. J.
(1969) Subjective probability in choice-reaction experiments, Journal of Mathematical Psychology, Vol. 6, pp. 81 120.

Lancaster, H. O.
(1958) The structure of bivariate distributions, Annals of Mathematical Statistics, Vol. 29, pp. 719 - 736.

Lance, G. N. & Williams, W. T.
(1966) Computer programs for hierarchical polythetic classification ("similarity analyses"), Computer Journal, Vol. 9, pp. 60 - 64.

Lazarsfeld, P. F.
(1950) The logical and mathematical foundations of latent structure analysis ; Chapter 10, Measurement and Prediction,S. A. Stouffer et al. (Eds.) Princeton Univ. Press.
(1954) A conceptual introduction to latent structure analysis; Chapter 7, Mathematical Thinking in the Social Sciences,P. F. Lazarsfeld (Ed.), The Free Press.

Leontief, W. W.
(1941) The Structure of the American Economy; 1919 1939: An Empirical Application of Equilibrium Analysis, N. Y.: Oxford Univ. Press.

Lewis, P. M. II.
(1959) Approximating probability distributions to redence storage requirements. Information and Control, Vol. 2, pp. 214 - 225.

Linfoot, E. H.
(1957) An information measure of correlation,Information and Control, Vol. 1, pp. 85 89.

Luce, R. D.
(1950) Connectivity and gnenralized cliques in sociometric group structure, Psychometrika, Vol. 15, pp. 169 - 190.
(1959) Individual Choice Behavior: A Theoretical Analysis, Wiley, N.Y.

Luce, R. D., Bush, R. R. & Galanter, E. (Eds.)
(1965) Handbook of Mathematical Psychology, Vol. I III.

Makino, H.
(1972) Methods and knacks to combine various climical tests (Rinshō kensa no kumiawase to sono kotsu), Kanehara Pub. Compony, (in Japanese).

Marschak, J. & Miyasawa, K.
(1968) Economic comparability of information systems, International Journal of Economic Review, Vol. 9, No. 2, pp. 137 - 174.

Max, J.
(1960) Quantizing for minimum distortion, IRE Trans. on Information Theory, Vol. 6, pp. 7 12.

Menger, K.
(1974) Morality, Decision and Social Organization: Toward a Logic of Ethics, D. Reidel, Dordrecht(Holland).

Miyasawa, K.
(1971) Introduction to information and decision theory (Jouhou kettei riron josetsu), Iwanami, (in Japanese).

Moreno, J. L.
(1934) Who shall survive?,Washingron, D. C., Nervous and Mental Disease Monograph No. 58.
(1960) The sociometry reader, Free Press.

Nemetz, T.
(1972) On the experimental determination of the entropy, Kybernetik,Bd. 10, Heft. 3, pp. 137 139.

Newcomb, T. M.
(1961) The Acquaintance Process. New York: Holt, Rinehalt and Winston.

Nishida, H. & Ara, M.
(1976) Theories and Techniques in Social Researches (Shakai chōsa no riron to giho), Vol. II, Kawashima Publ. Comp., Tokyo, (in Japanese).

Nishisato, S.
(1975) Practical Scaling Methods in Psychology (Ōyō shinri shakudo Kōsei hō) Seishin Pub. Comp., Tokyo, (in Japanese).
(1977) Recent developments in scaling and related areas: a bibliographic review (1), Japanese Jounal of Behaviormetrics (Published from the Behaviormetric Society of Japan), Vol. 4, No. 2, pp. 74 - 95, (in Japanese)
(1978) Recent developments in scaling and related areas: a bibliographic review (2), Japanese Journal of Behaviormetrics, Vol. 5, No. 2, pp. 37 55, (in Japanese).

Noguchi, S., Tomita, S. & Watanabe, T.
(1972) Relationship between organizing capability and system information utilized for system organization, pp. 391 413, in S.Watanabe(Ed.), Frontiers of Pattern Recognition, Academic Press.

Norman, M. F.
(1970) Limit theorems for additive learning models, Journal of Mathematical Psychology, Vol. 7, pp. 1 11.

Norman, R. Z.
(1972) A Derivation of a Measure of Relative Balance for Social Structures and a Characterization of Extensive Ratio Systems, Journal of Mathematical Psychology, Vol. 9, pp. 66 - 91.

Okuno, T., Kume, H., Haga, T. & Yoshizawa, T.
(1971) Methods of Multivariate Analysis (Tahenryo kaiseki hō), Nikkagiren, (in Japanese).

Ore, O.
(1962) Theory of Graphs, Amer. Math. Soc. Colloq. Publ., Vol. 38, Providence.

Pfaffelhuber, E.
(1971) Error estimation for the determination of entropy and information rate from relative frequencies, Kybernetik, Bd. 8, Heft. 2, pp.50 - 51.

Rajski, C.
(1961) A metric space of discrete probability distributions, Information and Control, Vol. 4, No. 4, pp. 371 377.

Rényi, A.
(1959a) New version of the probabilistic generalization of the large seive, Acta Math. Acad. Sci. Hung, Vol. 10, pp. 218 - 226.
(1959b) On measures of dependence, Acta Math. Acad. Sci. Hung, Vol. 10, pp. 441 - 451.

Rao, C. R.
(1952) Advanced Statistical Method in Biometric Research, Wiley, N. Y.

Rathie, P. N. & Kannappan, P.
(1972) A directed-divergence function of type β, Information and Control, Vol. 20, No. 1, pp. 38 45.

Romanovsky, V. I.
(1970) Discrete Markov Chains, Wolters-Noordhoff Publishing Groningen.

Romney, A. K., Shepard, R. N. & Nerlove, S. B.
(1972) Multidimensional Scaling: Theory and Applications in the Behavioral Science, Vol. II, Seminor Press, N. Y.

Ross, I. C. & Harary, F.
(1955) Identification of the liason persons of an organization using the structure matrix, Management Science, Vol. 1, pp. 251 258.

Rozeboom, W. W.
(1968) The theory of abstract partials: an introduction,Psychometrika, Vol. 33, No. 2, pp. 133 167.

Saito, T., Ogawa, S. & Nojima, E.
(1972a) Data analysis (1): summary review article on unidimensional scaling, Soken Kiyo (Reports of Research Activities on Computer Sciences and Technology, published from Nippon Univac Sogo Kenkyuso, Inc.), Vol. 2, No. 1, pp. 23 140, (in Japanese).
(1972b) Data analysis (2): summary review article on quantification theory, Soken Kiyo, Vol. 2, No. 2, pp. 17 - 226, (in Japanese).
(1973) Data analysis (3): summary review article on multidimensional scaling, Soken Kiyo, Vol. 3, No. 1, pp. 53 - 216, (in japanese).

Shannon, C. E. & Weaver, W.
(1949) The Mathematical Theory of Communications, Univ. of Illinios Press.

Shepard, R. N.
(1962a) The analysis of proximities: multidimensional scaling with an unknown distance function.I, Psychometrika, Vol. 27, No. 2, pp. 125 140.
(1962b) The analysis of proximities: multidimensional scaling with an unknown distance function. II,Psychometrika, Vol. 27, NO. 3, pp. 219 246.
(1966) Metric structures in original data, Jounal of Mathematical Psychology, Vol. 3, pp. 287 - 315.

Shepp, L. A.
(1965) Distinguishing a sequence of random variables from a translate of itself, Annals of Mathematical Statistics, Vol. 36, No. 4, pp. 1107 -1112.

Sibson, R.
(1969) Information radius,Z. Wahrscheinlichkeitstheorie verw, Geb. 14, pp. 149 160.

Simon, G.
(1973) Additivity of information in exponential family probability laws, Journal of the American Statistical Association, Vol. 68, No. 342, pp. 478 482.

Simons, G.
(1967) Lower bounds for average sample number of sequential multihypothesis tests, Annals of Mathematical Statistics, Vol. 38, No. 5, pp. 1343 1364.

Sokal, R. R. & Sneath, P. H.
(1963) Principles of Numerical Taxonomy, Freeman, W. H., San Francisco and London.

Solomon, R. L. & Wynne, L. C.
(1953) Traumatic avoidance learning: acquisition in normal dogs, Psychological Monograph, Vol. 67, No. 4, (Whole No. 354).

Spearman, C.
(1904) General intelligence, objectively determined and measured, American Journal of Psychology, Vol. 15, pp. 201 293.

Srikantan, K. S.
(1970) Canonical association between nominal measurements, Jounal of the American Statistical Association, Vol. 65, No. 329, pp. 284 292.

Sternberg, S.
(1963) Stochastic learning theory, In R. D. Luce, R. R. Bush, and E. Galanter(Eds.), Handbook of Mathematical Psychology, Vol. II, N. Y.: Wiley, pp. 1 120.

Stevens, S. S.
(1951) Mathematics, measurement and psychophysics, In S. S. Stevens(Ed.), Handbook of experimental psychology, Wiley, N. Y., pp. 1 - 49.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A. & Clausen, J. A.
(1950) Measurement and Prediction, Princeton Univ. Press, Princeton, N. J..

Takagi, S., Katai, O., Iwai, S., Imanaga, T., Endo, J. & Fujii, M.
(1972) Information theoretical considerations on pattern recognition of blood chemical components, the 15th Joint Convention on Automatic Control held by JAACE, SICE et al., pp. 343 - 344, (in Japanese).

Takagi, T. (Ed.)
(1971) Studies on Quantification in Psychology (Shinrigaku ni okeru sūryōka) Tokyo Univ. Press, Tokyo, (in Japanese).
(1972) Modern Psychology and Quantification (Gendai shinrigaku to sūryōka) Tokyo Univ. Press, Tokyo, (in Japanese).

Takane, Y.
(1977) On the relations among four methods of multidimensional scaling, Behaviormetrika (published from the Behaviormetric Society of Japan), No. 4, pp. 29 - 43.

Tallman, I.
(1967) The balance principle and normative discrepancy, Human Relations, Vol. 20, pp. 341 355.

Tanaka, K.
(1975) Theories and methods of sociometry (Soshiometori no riron to hōhō) Meiji Tosho, Tokyo, (in Japanese).

Tanaka, Y.
(1973) Scaling (Shakudo Kousei), Series on research methods in psychology, Vol. 16, Tokyo Univ. Press, Tokyo, (in Japanese).

Thomas, M. U. & Barr, D. R.
(1977) An Approximate Test of Markov Chain Lumpability, Jounal of the American Statistical Association, Vol. 72, No. 357, pp. 175 179.

Thurstone, L. L. & Jones, L. V.
(1957) The rational origin for measuring social values, Journal of American Statistical Association, Vol. 52, pp. 458 471.

Torgarson, W. S.
(1952) Multidimensional scaling: I. theory and method, Psychometrika, Vol. 17, No.4, pp. 401 - 419.

Toussaint, G. T.
(1971) Some upper bounds on error probability for multiclass pattern recognition, IEEE Trans. on Computers, Vol. C - 20, No. 8, pp. 943 - 944.
(1972) Some inequalities between distance measures for feature evaluation, IEEE Trans. on Computers, Vol. C - 21, No. 4, pp. 409 - 410.

Tsetlin, M. L.
(1961) On the behavior of finite automata in random media, Automatika i Telemekhanika, Vol. 22, pp. 1345 - 1354.
(1973) Automata Theory and Modeling of Biological Systems, N. Y.: Academic Press, pp. 12 18.

Varshavskii, V. I. & Vorontsova, I. P.
(1963) On the behavior of stochastic antomata with a variable structure, Automatika i Telemekhanika, Vol. 24, No. 3, pp. 353 - 360.

Vilmansen, T. R.
(1973) Feature evaluation with measures of probabilistic dependence, IEEE Trans. on Computers, Vol. C 22, No. 4, pp. 381 - 388.

von Wright, G. E.
(1968) An Essay in Deontic Logic and the General Theory of Action, North-Holland, Amsterdam.

Wallace, C. S. & Boulton, D. M.
(1968) An information measure for classification, Computer Journal,Vol. 11, pp. 185 194.

Williams, E. J.
(1959) Regression Analysis, Wiley, N. Y.

Winkler, R. L. & Murphy, A. H.
(1973) Information aggregation in probabilistic prediction, IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC 3, No. 2, pp. 154 160.

Wood, R. C.
(1969) On optimum quantization, IEEE Trans. on Information Theory, Vol. IT 15, No. 2, pp. 248 - 252.

Yasuda, S.
(1975) Social Statistics(Shakai tōkeigaku), Maruzen comp., Tokyo, (in Japanese)

Young, T. Y.
(1971) The reliability of linear feature extractors, IEEE Trans. on Computers, Vol. C - 20, No. 9, pp. 967 - 971.

Errata

		(original)	(revised)
Pi	L1B	aughor	author
P11	L1	An	A
	L20	*An*	*A*
P16	L13	strata	stratum
P31	L9	*quqntified*	*quantified*
P32	L9B	calculated	calculated.
P34	L7	$S(X_i)$	$S_R(X_i)$
P52	L15		$\triangleq$
	L5B	$\triangleq$	
P65	L5B	concerned	concerned.
P69	L20	Boolear	Boolean
P70	L5B	h'	$\mathbf{h}'$
P73	L9	that	the
P75	L10	otherhand	other hand
P76	L4	independece	independence
P86	L8	approximted	approximated
P87	L6B	evade	avoid
P90	L19	expericnce	experience
P97	L7B	$j(\bar{R})$	$J(\bar{R})$
P98	L1	fhown	shown
	L4	occures	occurs
P105	L7B	or	of
P111	L21	charactirized	characterized
P113	L6	of	of ρ
P114	L15	more	More
P148	L8B	as	
P149	L12	shawarz	Shawarz
P151	L5	meanigless	meaningless
P168	L2	to	of
P169	L17	It	
P181	L5B	classifing	classifying
P190	L4	ρ_i	ρ_i^*
	L5	ρ_i	ρ_i^*
P193	L7	tranposed	transposed
	L8B	minimal	minimum
P195	L7B	s_i	$s_i^!$
	L3B	is	is as
P203	L2	v_0	v_o
P210	L15	be-	bet-
	L27	illusrates	illustrates
P220	L2B	11.	11,
P227	L14	derect	direct
P237	L8B	set	set.
P240	L8	sufficience	sufficiency
P245	L8B	R_{i_j}	R_{ij}